

Prozedur und Algorithmen zur datenbasierten, anlagenweiten Regelkreisdiagnostik

Von der Fakultät für Ingenieurwissenschaften der
Universität Duisburg-Essen
Abteilung Elektrotechnik und Informationstechnik

zur Erlangung des akademischen Grades eines

Doktor-Ingenieurs (Dr.-Ing.)

genehmigte Dissertation

von

Markus Stockmann
aus
Bergisch Gladbach

1. Gutachter: Univ.-Prof. Dr.-Ing. Steven X. Ding
 2. Gutachter: Univ.-Prof. Dr.-Ing. Dirk Söffker
- Tag der mündlichen Prüfung: 11.10.2012

Danksagung zur kooperativen Promotion

Die vorliegende Arbeit entstand während meiner Tätigkeit als wissenschaftlicher Mitarbeiter im Institut Anlagen- und Verfahrenstechnik der Fakultät für Anlagen, Energie- und Maschinensysteme der Fachhochschule Köln. Für die finanzielle Unterstützung während der Forschungszeit möchte ich mich bei den Verantwortlichen der Kompetenzplattform STEPS (Sustainable Technologies and Computational Services for Environmental and Production Processes) der Fachhochschule Köln und des Landes Nordrhein-Westfalen bedanken.

Des Weiteren schulde ich meinem Betreuer von Seiten der Fachhochschule Köln, Herrn Professor Robert Haber, Dank, der mich bereits seit vielen Jahren auf diese Arbeit vorbereitet hat. Sein Lob, seine Erfahrungen und unsere regen Diskussionen tragen unschätzbaren Anteil an der vorliegenden Arbeit.

Ich möchte mich genau so herzlich bei meinem Doktorvater, Herrn Professor Steven X. Ding, bedanken, der mir das kooperative Promotionsverfahren zwischen der Universität Duisburg-Essen und der Fachhochschule Köln erst ermöglicht hat. Ohne seine wissenschaftliche Kritik und seine Anregungen wäre diese Arbeit in dieser Form niemals zustande gekommen.

Private Danksagung

Natürlich möchte ich mich ebenfalls bei meiner Frau Lena bedanken, die durch unzählig' aufbauende Worte und liebevolle Unterstützungen so viel dazu beigetragen hat. Zum Schluss geht mein Dank selbstverständlich auch an meine Mutter, die mir ein Studium und somit auch die Promotion erst ermöglicht hat, trotz so mancher Steine im Weg.

Vielen Dank!

Inhalt

1 Einleitung	9
1.1 Regelungsqualität und Begriffserklärungen	10
1.2 Industrielle Anforderungen	11
1.3 Motivation und Zielsetzung	13
1.4 Organisation der Kapitel	14
2 Erkennung anlagenweiter Fehlerverursacher	16
2.1 Problemstellung	17
2.2 Erkennung im Offline-Betrieb	18
2.2.1 Erkennung im Falle von ausschließlich linearen Teilprozessen	22
2.2.2 Erweiterter Ansatz für nichtlineare Systeme und MISO-Systeme	26
2.2.3 Simulationsergebnisse	39
2.2.4 Laufzeitanalyse	44
2.3 Vereinfachte Erkennung im Online-Betrieb	46
2.3.1 Grundlagen	46
2.3.2 Simulationsergebnisse	48
2.3.3 Schätzung des Spektrums mit geringer Rechenleistung	49
2.3.4 Online-Einsatz	51
2.3.5 Laufzeitanalyse	51
2.4 Zusammenfassende Bemerkung	52
3 Fehlererkennung einzelner Regelkreise	54
3.1 Fehlererkennung mittels Performance-Indizes	56
3.1.1 Harris-Index	56
3.1.2 Vorhersage des Regelfehlers	60
3.1.3 Berechnung individueller Maßzahlen	61
3.2 Fehlererkennung mittels Oszillationserkennung	62
4 Allgemeine Fehlerdiagnose einzelner Regelkreise	67
4.1 Problemstellung	67
4.2 Beschreibung der gängigsten Regelungsszenarien	68
4.3 Beschreibung des Regelungsszenarios Haft- und Gleitreibung	74
4.4 Automatisierte Fehlerdiagnose einzelner Regelkreise	81
4.5 Zusammenfassende Bemerkung	96
5 Validierung und Testen an Benchmark-Anlagen	99
5.1 Simulierte Benchmarks	99
5.2 Industrielle Benchmarks	107
6 Zusammenfassung	111

Abbildungsverzeichnis

Abbildung 1.1: Optimale Einbindung des anlagenweiten Diagnosesystem	13
Abbildung 1.2: Organisation der Kapitel	15
Abbildung 2.1: Abstrahierte Anlage mit M Teilprozessen und Kreislaufführung	17
Abbildung 2.2: Prozedur zur Identifikation von anlagenweiten Fehlerverursachern	20
Abbildung 2.3: Kreuzkorrelationsfunktion zwischen S_i und S_j , Eingang S_i und S_j	23
Abbildung 2.4: Kreuzkorrelationsfunktion zwischen S_i und S_j , Eingang S_i	24
Abbildung 2.5: Anlagenmodell mit $M = 3$ linearen Teilprozessen	24
Abbildung 2.6: Visualisierung der Punktmengen	27
Abbildung 2.7: Differenz zwischen \hat{y}_{k_1+nd} und y_{k_0+nd} abhängig von der Verschiebung	29
Abbildung 2.8: Ersetzung von fehlenden Messwerten mittels k nächster Nachbarn ($k = 1$) ...	29
Abbildung 2.9: Prozesstypen zur Totzeitschätzung	33
Abbildung 2.10: Zeitverlauf der vier Prozesstypen	33
Abbildung 2.11: Totzeitschätzung für MISO-Prozess	34
Abbildung 2.12: Einfluss der Anzahl an Messungen und der nächsten Nachbarn	36
Abbildung 2.13: Totzeitschätzung für Prozesse	38
Abbildung 2.14: Totzeitschätzung für dynamische Prozesse	38
Abbildung 2.15: Totzeitschätzung für dynamische Prozesse mit $p = 1$ und $p = 4$	39
Abbildung 2.16: Benchmarkprozess zur Identifikation von Fehlerverursachern	39
Abbildung 2.17: Skalierte Zeitverläufe im Falle der normalen Betriebsweise	40
Abbildung 2.18: Skalierte Zeitverläufe im Falle einer Oszillation in S_3	40
Abbildung 2.19: Skalierte Zeitverläufe im Falle einer Oszillation in S_3	41
Abbildung 2.20: Skalierte Zeitverläufe im Falle einer arbiträren Störung in S_2	42
Abbildung 2.21: Skalierte Zeitverläufe im Falle einer Oszillation in S_1	42
Abbildung 2.22: Skalierte Zeitverläufe im Falle einer Oszillation und externer Störung	43
Abbildung 2.23: Skalierte Zeitverläufe im Falle einer Oszillation in S_4	44
Abbildung 2.24: Rechenzeit in Abhängigkeit der Vektorlänge	45
Abbildung 2.25: Amplitudenverteilung	47
Abbildung 2.26: Resonanzspektrum S_1	48
Abbildung 2.27: Schema zur Schätzung des Spektrums	49
Abbildung 2.28: FFT-Spektrum und geschätztes Spektrum	50
Abbildung 2.29: Kritische Amplituden zur Erkennung von signifikanten Amplituden	51
Abbildung 2.30: Benutzeroberfläche des Programms zur Verursachererkennung	51
Abbildung 2.31: Rechenzeit des Online-Verfahrens	52
Abbildung 2.32: Fehlerfortpflanzungsmodell im Falle keines anlagenweiten Fehlers	53
Abbildung 2.33: Fehlerfortpflanzungsmodell im Falle eines anlagenweiten Fehlers	53
Abbildung 3.1: Prozess zur Wirksamkeitseinschätzung der Fehlererkennungsmethoden	55
Abbildung 3.2: Verschiedene Regelungsszenarien nach Sollwertsprung von 0 auf 1	55
Abbildung 3.3: Regelgrößenverlauf bei gut eingestelltem Regler	57
Abbildung 3.4: Autokorrelationsfunktion bei gut eingestelltem Regler mit der Totzeit 5	58
Abbildung 3.5: Skalierter Regelgrößenverlauf bei aggressiv eingestelltem Regler	58
Abbildung 3.6: Autokorrelationsverläufe für verschiedene Regelungsszenarien	59
Abbildung 3.7: Verschiedene Regelungsszenarien zum Testen der Fehlererkennung	61
Abbildung 3.8: Signal mit dargestellten Nulldurchgängen	63

Abbildung 3.9: Autokorrelationsfunktion eines Signals mit Nulldurchgängen	64
Abbildung 3.10: Häufigkeiten als Ergebnis der Monte-Carlo Sim. bei 5000 Realisierungen ..	64
Abbildung 3.11: Häufigkeiten als Ergebnis der Monte-Carlo Sim. bei 100 Realisierungen ...	64
Abbildung 4.1: Akzeptable und aggressive Regelung	68
Abbildung 4.2: Aggressive Industrieregung	69
Abbildung 4.3: Industrieregung mit oszillierendem Sollwert	70
Abbildung 4.4: Einfluss des Reglertyps (PI und PID) auf die Stellgeräteaktivität	70
Abbildung 4.5: Begrenzte Industrieregung	71
Abbildung 4.6: Inaktive Industrieregung in Folge einer instabilen Regelung	72
Abbildung 4.7: Gleichgerichtetheit einer langsamen und akzeptablen Regelung	72
Abbildung 4.8: Industrieregung mit einer externen Störung	73
Abbildung 4.9: Darstellung eines pneumatischen Ventils	74
Abbildung 4.10: Darstellung des Totbands bei Zahnradantrieb	75
Abbildung 4.11: Ein- und Ausgangsverhalten bei Stiction Modell	76
Abbildung 4.12: Stictionverhalten von Stellsignal, Ventilöffnung und Regelgröße	77
Abbildung 4.13: Schritte der Rohdatenvorbehandlung	78
Abbildung 4.14: Trainingsdaten im Raum der ersten beiden Hauptkomponenten	79
Abbildung 4.15: Simuliertes Muster mit unbekannter Zugehörigkeit	79
Abbildung 4.16: Industrielles Muster mit unbekannter Zugehörigkeit	80
Abbildung 4.17: Verschiedene Rauschintensitäten bei verschiedenen Sollwerten	84
Abbildung 4.18: Dreimalige Differenzierung eines geregelten Prozesses	85
Abbildung 4.19: Berechnete p_4 Werte für 20 Industriedatensätze	86
Abbildung 4.20: Stellsignalverlauf hervorgerufen durch Haft- und Gleitreibung	86
Abbildung 4.21: Histogramme für idealen dreiecksförmigen und sinusförmigen Verlauf	87
Abbildung 4.22: Histogramme für gestörten dreiecksförmigen und sinusförmigen Verlauf ..	88
Abbildung 4.23: Kumulierte Häufigkeiten gegenüber dem Klassenindex	89
Abbildung 4.24: Relative Häufigkeiten für Dreiecksverlauf	87
Abbildung 4.25: Relative Häufigkeiten für Sinusverlauf	87
Abbildung 4.26: Monte-Carlo basierte Verteilung der berechneten p_6 Werte	92
Abbildung 4.27: Verschiedene deterministische Störungen	93
Abbildung 4.28 Verschiedene deterministische Störungen mit Ausreißererkennung	94
Abbildung 5.1: Vereinfachtes Verfahrensfließbild einer industriellen Hydrocrackeranlage.	100
Abbildung 5.2: Modell einer simulierten Hydrocrackeranlage	100
Abbildung 5.3: Modell eines simulierten katalytischen Festbetts	101
Abbildung 5.4: Modell einer simulierten Quenches	101
Abbildung 5.5: Daten aller Messungen und simulierter Haft- und Gleitreibung	101
Abbildung 5.6: Spektrum des verursachenden Regelkreises	103
Abbildung 5.7: Spektren aller Teilprozesse bei Verursacher S_6	103
Abbildung 5.8: Daten aller Messungen bei oszillierendem Sollwert in S_{11}	105
Abbildung 5.9: Spektren aller Teilprozesse bei Verursacher S_{11}	106
Abbildung 5.10: Zeitlicher Verlauf aller relevanten Hydrocrackermessungen	107
Abbildung 5.11: Spektren aller Teilprozesse bei unbekanntem Verursacher	109
Abbildung 5.12: Wechsel von Kaskadenregelung auf manuelle Regelung	110

Tabellenverzeichnis

Tabelle 2.1: Kreuzkorrelationsfunktionen aller Permutationen	25
Tabelle 2.2: Relative Störungsenergien bei der Störungsfrequenz 0.1 und Amplitude 2	49
Tabelle 2.3: Relative Störungsenergien bei der Störungsfrequenz 0.2 und Amplitude 1	49
Tabelle 3.1: c_{lp} -Werte der verschiedenen Regelungsszenarien	62
Tabelle 3.2: n_{oi} Werte der verschiedenen Regelungsszenarien	65
Tabelle 4.1: Erkennungsrate bei verschiedenen Prozessen	80
Tabelle 4.2: Regelungsszenarien mit den entsprechenden Charakteristika	81
Tabelle 4.3: Zusammenhang zwischen Regelungssituation und Performance-Indizes	96
Tabelle 4.4: Berechnete Performance-Indizes der Benchmarkprozesse	97
Tabelle 4.5: Berechnete Performance-Indizes der Benchmarkprozesse	97
Tabelle 5.1: Berechnete Performance-Indizes bei Haft- und Gleitreibung in S_6	104
Tabelle 5.2: Berechnete Performance-Indizes bei schwingendem Sollwert in S_{11}	107
Tabelle 5.3: Berechnete Performance-Indizes des industriellen Prozesses.....	110

Symbolverzeichnis

Mathematische Variablen und Symbole

\hat{a}	Schätzung von a
$a \in [a_{\min}, a_{\max}]$	a sei Element der Menge von inklusive a_{\min} bis exklusive a_{\max}
$b = \arg \max_{b \in [c, d]} (f(a))$	$f(b)$ ist Maximum der Funktion $f(a)$ im Bereich c bis d
$d((a_1, b_1), (a_2, b_2))$	Euklidische Distanz zwischen zwei Punkten
\in	Element
$\text{sign}(a)$	Vorzeichen von a mit den möglichen Werten -1, 0 und 1
$x^{(i)}$	i -fach diskret differenziertes Signal von x
\bar{z}	Komplex Konjugiertes von z

Regelungstechnische Variablen und Symbole

d	Diskrete Totzeit, diskrete Zeitverschiebung ($T_d / \Delta T$)
$d_{i,j}$	Zeitverzug (Totzeit) zwischem i -ten und j -ten Teilprozess
$e_k, e(k),$ $e_b, e(t)$	Reglerabweichung ($w_k - y_k$) eines Reglers zum k -ten Abtastpunkt
$\varepsilon_k, \varepsilon(k),$ $\varepsilon_b, \varepsilon(t)$	Rauschen und Störungen zum k -ten Abtastpunkt
$G_s(s), G_p(s)$	Übertragungsfunktion eines Prozesses $G_p(s) = G_s(s) = \frac{Y(s)}{U(s)}$, Beispiel Prozess erster Ordnung mit Totzeit: $T_1 = 0.75, K_{PS} = 1, d = 3, \Delta T = 0.05$ $\Rightarrow G_p(s) = \frac{1}{0.75s + 1} \cdot e^{-s \cdot 0.15}$. Zur Verdeutlichung wird die Abtastzeit stets mit angegeben
k	Diskrete Zeit (Abtastpunkt)
K_{PR}, T_N, T_V	Reglerverstärker (<u>P</u> roportionaler Anteil), Nachstellzeit (<u>I</u> ntegrierender Anteil) und Vorhaltezeit (<u>D</u> ifferenzierender Anteil) eines (P)(I)(D)- Reglers der Form $u(t) = K_{PR} \cdot \left(e(t) + T_N^{-1} \cdot \int_0^t e(\tau) d\tau + T_V \cdot de(t)/dt \right)$
K_{PS}	Prozessverstärkung
M_u	Anzahl an Systemeingängen

q^{-n}	Verschiebeoperator, $q^{-n} := y_{k-n}, n=1,2,\dots$.Relation zur z -Transformation: $y_k + \frac{3}{2}y_{k-1} + \frac{1}{2}y_{k-2} = \frac{1}{2}u_k + \frac{1}{2}u_{k-1} - \frac{1}{2}u_{k-2}$ $\Rightarrow y_k + \frac{3}{2}q^{-1}y_k + \frac{1}{2}q^{-2}y_k = \frac{1}{2}u_k + \frac{1}{2}q^{-1}u_k - \frac{1}{2}q^{-2}u_k$ $\Leftrightarrow \frac{y_k}{u_k} = \frac{\frac{1}{2} + \frac{1}{2}q^{-1} - \frac{1}{2}q^{-2}}{1 + \frac{3}{2}q^{-1} + \frac{1}{2}q^{-2}} = \frac{z^2 + z - 1}{2z^2 + 3z + 1} = \frac{Y(z)}{U(z)}$
S_i	Statischer Teilprozess der beispielhaften Form $y_i = S_i(u_i)$
s_i	Messwerte des i -ten Teilprozess
t, τ	Kontinuierliche Zeit
T_d	Totzeit
ΔT	Abtastzeit
$w_k, w(k),$	Sollwert eines Reglers zum k -ten Abtastpunkt
$w_b, w(t)$	
$y_k, y(k),$	Regelgröße eines Reglers zum k -ten Abtastpunkt
$y_b, y(t)$	
x^*	Skalierter (normierter) Datenvektor x

Statistische Variablen und Symbole

$acf_u(k)$	Wert der Kreuzkorrelationsfunktion von u bei Verschiebung k
$ccf_{u,y}(k)$	Wert der Kreuzkorrelationsfunktion zwischen u und y bei Verschiebung k
$Cov(X,Y)$	Kovarianz zwischen X und Y
$E(X), \mu_X$	Erwartungswert der Zufallsvariablen X
$h(u,y)$	Shannon-Entropie zwischen y und u
$MAD(X)$	Median der absoluten Medianabweichung der Zufallsvariablen X
$MSE(X)$	Mittlere quadratische Abweichung von X
$p(a)$	Wahrscheinlichkeit für das Auftreten von a
$p(a,b)$	Wahrscheinlichkeit für das gleichzeitige Auftreten von a und b
$p(a/b)$	Wahrscheinlichkeit für das Auftreten von a unter der Bedingung, dass b bereits aufgetreten ist.
$t_{u \rightarrow y}(u, y),$	Richtungsabhängige Transferentropie zwischen u und y
$t_{y \rightarrow u}(y, u)$	
$Var(X), \sigma_X^2,$	Varianz der Zufallsvariablen X
$\sigma^2(X)$	
X, Y	(Mengen der) Zufallsvariablen X und Y
$X \sim \exp(\lambda)$	Exponentialverteilung von X mit Verteilungsparameter λ
$X \sim N(\mu_X, \sigma_X^2)$	Normalverteilung von X
$X \sim \chi^2(\mu_{\chi^2})$	Chi-Quadratverteilung von mit Verteilungsparameter μ_{χ^2}
$X_{0.5}$	Median der Zufallsvariablen X

Sonstige Variablen und Symbole

a_{sin}	Amplitude einer sinusförmigen Schwingung
a_0, a_1, \dots	(Schätz-) Parameter für polynomische Funktion oder Regression,
b_0, b_1, \dots	bzw. ARX (Auto Regression und externe Anregung) Parameter
α	Irrtumswahrscheinlichkeit
b	Achsenabschnitt einer linearisierten Verteilungsfunktion, Horizont bei Vorhersage des Regelfehlers
c_{lp}	Individuelle Maßzahl zur Abschätzung der Regelkreisperformance
$e_{dist}(f_1)$	Übertragene Leistung im Bereich von $\Delta f_k + \Delta f_g$ um Frequenz f_1
e_{ges}	Gesamt übertragene Leistung basierend auf Frequenzspektrum
$e_{sist,i}$	Boolsche Variable, ob die Regelgröße an Abtastpunkt i stationär und betragsmäßig größer als Grenzwert ist.
ε_{eig}	Verhältnis der Eigenwerte zur Bestimmung der Linearität
$\varepsilon_{out,k}$	Boolsche Variable, ob ein differenzierter Regelfehler zum Abtastpunkt k signifikant ist
f	Frequenz
$f_{nyquist}$	Nyquist Frequenz (=halbe Abtastfrequenz)
f_p	Individuelle Maßzahl zur Einschätzung der Fehlerfortpflanzung eines einzelnen Regelkreises
f_1, f_2, \dots	Impulsantwortkoeffizienten
h_u	Empirische Klassenhäufigkeiten
h_u^{kum}	Kumulierte empirische Klassenhäufigkeiten
θ_i	Relative Position des i -ten Maximums einer Halbperiode
i^+, i^-	Idle Index bei Prozess mit positiver ($^+$) und negativer ($^-$) Prozessverstärkung
J	Simulationsparameter zur Ausprägtheit von Haft- und Gleitreibung in Ventilkörpern
k_{imp}	Anzahl an nächsten Nachbarn bei k_{imp} -nächste-Nachbarn-Ersetzung
$k_{0max,i}$	Abtastpunkt mit maximalem absoluten Signalwert zwischen zwei Nulldurchgängen
$\Delta k_{0,i}$	Abtastpunkte zwischen i -tem und $i+1$ -tem Nulldurchgang
κ	Angenommene diskrete Verschiebung zweier Signale
l	Anzahl der Nulldurchgänge, Länge Vektor bei Transferentropie
m	Steigung und Achsenabschnitt der linearisierten Verteilungsfunktion, Länge Vektor bei Transferentropie
M	Anzahl an Teilprozessen in Anlage
n	Anzahl an Messwerten
n_d	Angenommene diskrete Verschiebung zweier Signale
n_k	Anzahl an Histogrammklassen
n_{oi}	Normalisierter Oszilationsindex
n_s	Anzahl an interpolierten Splinewerten
η	Diverse Wirkungsgrade im Bereich von 0 bis 1, z.B. Harris-Index
	η_{Harris}
$O(\cdot)$	Obere Laufzeitklasse, abhängig von Datenmenge
$P(x_i)$	Punktmenge von x_i , mit $x_i \in \mathbf{R}^n$
p_1, \dots, p_{10}	Performance Zahlen zur automatisierten Regelkreisdiagnostik

p, q, r, s	Vereinfachende Substitutionen zur Berechnung der Verteilungsfunktion
p, q	Modellordnung
PC_i	Scores (Werte der linearen Abbildung) auf i -ter Hauptkomponente
p_{yy}	Quadratische Amplituden der Fouriertransformation
q_0	Regelmäßigkeit von Nulldurchgängen
r	Residuum zwischen geschätztem und tatsächlichem Wert
r_i	Relativer Klassenindex der i -ten Klasse
ρ_{crit}	Signifikanzschwelle von ρ
$\rho(n_d)$	Funktion zur Schätzung der Zeitverschiebung, basierend auf k_{imp} - nächster-Nachbar-Ersetzung
S	Simulationsparameter zur Ausgeprägtheit von Haft- und Gleitreibung in Ventilkörpern
$u_{i,k}$	i -ter Prozesseingang (ggf. angenommen) zu k -tem Abtastpunkt
$u_{ins,k}$	Instationäre Bereiche von u zu k -tem Abtastpunkt
u_{min}, u_{max}	Minimales und maximales Stellsignal (aus Daten geschätzt)
U_{min}, U_{max}	Minimales und maximales Stellsignal (a-priori bekannt)
w_1, w_2, \dots	Gewichtungsfaktoren
$y_{i,k}$	i -ter Prozessausgang (ggf. angenommen) zu k -tem Abtastpunkt
$y_{ins,k}$	Instationäre Bereiche von y zu k -tem Abtastpunkt
\mathbf{Z}	Beeinflussungsmatrix
$\mathbf{Z}_{ccf}, \mathbf{Z}_{knn}$	Beeinflussungsmatrix basierend auf nächste-Nachbarn-Ersetzung oder Kreuzkorrelationsfunktion

Konventionen

Zur besseren Verständlichkeit sind Zeiteinheiten standardmäßig nicht angegeben. Abtastzeit und Zeitkonstanten beziehen sich jedoch stets auf dieselbe verwendete Zeiteinheit. Außerdem soll gelten, dass Zählindizes (z.B. i, j, k, κ, n_d) und Operatoren (s, q, z) klein und kursiv dargestellt werden. Skalare werden in Anlehnung daran im Folgenden ebenfalls klein und kursiv in der Form x , ggf. mit einem weiteren Suffix in der Form x_y dargestellt, wobei y kein Zählindex ist und lediglich der besseren Unterscheidbarkeit oder der Zugehörigkeit dient. Vektoren werden in derselben Form dargestellt, wobei diese jedoch, je nach besserer Lesbarkeit, entweder einen Zählindex als Suffix enthalten, z.B. x_k oder den Zählindex in runden Klammern angeben, z.B. $x(k)$. Konstanten (im Folgenden per Definition nicht abhängig von der Analyse, z.B. Prozessparameter) werden in kursiver Großschreibweise dargestellt und beeinhaltet bei Verwechslungsgefahr mit anderen Konstanten häufig ein Suffix in Anlehnung an bekannte Konventionen (z.B. T_N, T_V). Matrizen und Mengen werden im Folgenden ebenfalls in Großschreibweise dargestellt, wobei Matrizen dickgedruckt dargestellt werden. Zur Kenntlichmachung von komplexen Zahlen werden diese in dickgedruckter Kleinschreibweise dargestellt. Ausnahmen von dieser Konvention werden zugelassen, wenn alternative Schreibweisen sich in der Fachliteratur durchgesetzt haben, z.B. $X_{0.5}$ für den Median oder d für die (konstante) diskrete Totzeit.

Kapitel 1

Einleitung

Seit der Einführung des Harris-Index [25] im Jahre 1989 hat die Entwicklung von Methoden zur Beurteilung der Regelgüte und die damit verbundene Optimierung einzelner Regler stetig zugenommen. Mit der gestiegenen Verfügbarkeit industrieller Rechner wurde dabei ein Hauptaugenmerk verstärkt auf den Online-Einsatz einer solchen Beurteilung und Optimierung gerichtet. Den Grund für die Notwendigkeit dieser Entwicklung liefert eine industrielle Studie [29] mit dem Ergebnis, dass 25% aller Regelungen im industriellen Umfeld inaktiv sind. Weitere 25% aller Regelungen führen zu einer Verschlechterung der Regelungssituation und lediglich die Hälfte aller eingesetzten Regelungen ist zufrieden stellend ausgelegt. Es besteht somit ein enormes Interesse vonseiten der Industrie an einer zeitnahen Beurteilung der Regelgüte und einer damit eventuell verbundenen Regleroptimierung, da schnellere Sollwertregelungen oder besser eingestellte Störgrößenregelungen häufig direkt zu einer Steigerung des Durchsatzes oder zu einer Verbesserung der Produktqualität führen. Was sich letztendlich in höheren Umsatzzahlen widerspiegelt.

Die rasant angestiegene Computerleistung der letzten Jahre hat anschließend dazu geführt, dass der eingeschränkte Blick auf das Reglertuning sich auf eine komplexere Regelkreisdiagnose erweitert hat. Diese ergänzt dabei die bekannte Reglerdiagnose um zusätzliche Untersuchungen des Prozesses (z.B. Nichtlinearitäten), der Sensorik (z.B. eingefrorene Messungen) und der Aktorik (z.B. defekte Regelventile). Mithilfe dieses erweiterten Ansatzes ist es möglich, gute Einschätzungen der gesamten Regelkreisqualität zu treffen und gleichzeitig Vorschläge zur Verbesserung zu liefern. Aufgrund seiner Komplexität ist der Schritt der Diagnose sehr rechenintensiv, laut [29] beträgt die Zeit für die Diagnose eines einzelnen Regelkreises zwischen vier und sechs Stunden. Hinzukommt, dass eine umfassende Diagnose im alltäglichen Gebrauch durchaus fehleranfällig ist. Sollen alle Regelkreise einer gesamten Anlage diagnostiziert werden, kommt es somit in der Regel zu einer Fülle von Fehldiagnosen.

Um die Zahl der fehlerhaften Diagnosen zu verringern und um Zeit zu sparen, wird seit einiger Zeit stark auf dem Gebiet der anlagenweiten Diagnostik geforscht. Die anlagenweite Diagnose wird eingesetzt, wenn einzelne Regelkreise mit einer minderen Güte die restlichen Regelkreise einer Anlage derart beeinflussen, dass diese ebenfalls eine geringe Performance aufweisen. Die einfache und klassische Regelkreisdiagnose bestünde in so einem Fall darin, jeden einzelnen Regelkreis zu diagnostizieren und unter der Annahme, dass es zu keinen Fehldiagnosen kommt, lediglich im verursachenden Regelkreis ein zugrunde liegendes Problem zu diagnostizieren. Die anlagenweite Regelkreisdiagnostik ergänzt den Schritt der einfachen Regelkreisdiagnose um eine

voran geschaltete Verursacheridentifizierung. Daran anschließend muss die einfache Regelkreisdiagnose lediglich in dem verursachenden Regelkreis durchgeführt werden.

Da es für die anlagenweite Regelkreisdiagnostik zum heutigen Zeitpunkt keine oder nur unzulängliche Online-Verfahren [12] gibt, beschäftigt sich die vorliegende Arbeit mit der Entwicklung eines solchen Gesamtkonzepts. Da, wie beschrieben, die einzelne Regelkreisdiagnostik ebenfalls Teil dieses Gesamtkonzepts sein muss, wird diese ebenfalls innerhalb dieser Arbeit dargestellt. Ein Hauptaugenmerk liegt bei dabei auf den industriellen Anforderungen an ein solches System. Diese werden neben den wichtigsten Begriffen in diesem Kapitel genannt.

1.1 Regelungsqualität und Begriffserklärungen

Die industriellen Anforderungen an Regelungen variieren stark je nach Einsatzgebiet; so müssen Regelungen von sicherheitsrelevanten Anlagenteilen stets sicherstellen, dass Störungen rasch ausgeregelt werden um den sicheren Arbeitspunkt nicht zu verlassen. Dies gilt z.B. in der Petrochemie bei der Regelung von Rektifikationskolonnen zur Herstellung von brennbaren Destillaten. Im Vergleich dazu liegt das Hauptaugenmerk der Kunststoffindustrie nicht so sehr auf der Ausregelgeschwindigkeit, sondern vielmehr auf der Tatsache, dass die zu verarbeitenden Polymere keinen starken Temperaturschwankungen ausgesetzt werden [46]. Eine allgemeine Definition von Regelungsqualität ist somit grundsätzlich unmöglich. Dennoch lassen sich in der Literatur einige Kriterien finden, die, für eine Vielzahl von Einschränkungen, zumindest eine grobe Abschätzung der Regelungsqualität liefern. Zu den wohl bekanntesten Kriterien zählt der Harris-Index nach [25], der CLPA-Index (*Control Loop Performance Assessment*, Regelkreis Performance Abschätzung) nach [11] oder die Verwendung individuell anpassbarer Maßzahlen, z.B. nach [61]. Für eine genaue Beschreibung dieser Kriterien wird auf Kapitel 3 verwiesen.

Auch wenn die Rahmenbedingungen für eine gute Regelung sehr schwierig aufzustellen sind, so lassen sich jedoch für nahezu alle Industriezweige vergleichbare Anzeichen einer schlechten Regelung definieren. Nach [10] sind dies die *Regelungsszenarien* „Prozessprobleme und Beschränkungen“, „Fehlerhafte Reglereinstellungen“ und „Probleme im Regelventil“. Da diese Anzeichen als zu ungenau und zu begrenzt angesehen werden, werden im Folgenden acht Regelungsszenarien vorgestellt, die als detaillierter angesehen werden können und somit als Anzeichen einer schlechten Regelungsqualität gedeutet werden sollen:

- aggressive Reglereinstellung,
- schwingender Sollwert,
- Haft- und Gleitreibung im Ventilkörper,
- externe Störungen,
- Messrauschen,
- Begrenzungen,
- inaktive Regelung,
- träge Reglereinstellung.

Eine gute Regelungsqualität zeichnet sich somit indirekt durch das Nichtvorhandensein der genannten Regelungsszenarien aus.

Jedes Regelungsszenario besitzt dabei bestimmte *Charakteristika*; so manifestiert sich eine aggressive Reglereinstellung häufig in einer oszillierenden Regelgröße, da der Regler an seiner Stabilitätsgrenze arbeitet. Die Menge aller Charakteristika wird in der vorliegenden Arbeit als *Fehler* verallgemeinert beschrieben.

Fehler: Durch ein Regelungsszenario hervorgerufene Fehlfunktion des Regelkreises.

In der vorliegenden Arbeit wird gezeigt, dass ein wichtiger Schritt der anlagenweiten Regelkreisdiagnose aus der Fehlererkennung besteht.

Fehlererkennung: Suche nach den typischen Charakteristika der beschriebenen Regelungsszenarien.

Ist ein Fehler erkannt, so kann davon ausgegangen werden, dass die Regelung eine unzufrieden stellende Regelungsqualität besitzt. Um jedoch Aussagen zu ermöglichen, wie die Qualität verbessert werden kann und um weitere Schwachstellen des Regelkreises aufzudecken, ist es wichtig, dass im Schritt der *Fehlerdiagnose*, das dem Fehler zugrunde liegende Szenario eindeutig erkannt wird.

Fehlerdiagnose: Bestimmung des Regelungsszenarios, welches zu dem erkannten Fehler geführt hat.

Führt der Fehler eines einzelnen Regelkreises dazu, dass alle anderen Regelkreise der Anlage denselben Fehler (meist in abgeschwächter Form) zeigen, so wird im Folgenden von einem anlagenweiten Fehler gesprochen.

Anlagenweiter Fehler: Alle Regelkreise einer Anlage zeigen einen bestimmten Fehler, zurückzuführen auf das Regelungsszenario eines einzelnen, beteiligten Regelkreises.

Nicht alle der genannten Szenarien führen immer zur Entstehung eines anlagenweiten Fehlers, so führt z.B. eine Beschränkung in einem Regelkreis nicht zwingend dazu, dass die nachfolgenden Regelkreise diesen Fehler auch zeigen. Eine Ausnahme hierfür ist allerdings, dass die folgenden Regler durch den begrenzten Regler ebenfalls an ihre eigenen Grenzen verlagert werden. Im Gegensatz dazu führt eine aggressive Regelung oder ein Ventil, welches starke Haft- und Gleitreibung zeigt, nahezu immer dazu, dass eine Oszillation als anlagenweiter Fehler entsteht.

1.2 Industrielle Anforderungen

Die industriellen Anforderungen an die (anlagenweite) Regelkreisdiagnostik sind enorm. So fasst [37] etwa die wichtigsten Anforderungen aus [30], [64] und [34] wie folgt zusammen:

- *Nicht invasiver Betrieb, ohne Anlagentests:* Die verwendete Diagnostik muss ohne Anlagentests auskommen und darf den laufenden Anlagenbetrieb nicht stören, auch wenn diese häufig bessere Ergebnisse liefern, als der geschlossene Betrieb, z.B. durch die Auswertung von Sprungantworten.
- *Voll automatisierbar:* Die Diagnostik soll nach Möglichkeit ohne Inanspruchnahme der Anlagenfahrer und Ingenieure arbeiten.

- *Nahezu fehlerfrei:* Fehllarme oder versäumte Alarme führen dazu, dass die Entscheidungen der Regelkreisdiagnostik ignoriert werden.
- *Verwendung von unbearbeiteten Betriebsdaten:* Die Verwendung von historischen Betriebsdaten, welche in der Regel komprimiert oder gefiltert sind, führt nach [59] und [13] zu fehlerhaften Einschätzungen.
- *Einfachheit der Ergebnisse:* Um die verantwortlichen Anlagenbetreiber nicht im Falle eines Alarms mit Informationen zu überfluten, sollten die Ergebnisse der Diagnostik möglichst simpel und überschaubar sein.
- *Priorisierung der erkannten Fehler:* Die erkannten Fehler sollen nach ihrer Relevanz sortiert werden [42].

Neben den genannten Anforderungen seien jedoch weitere abgeschwächte Wünsche aus der eigenen industriellen Erfahrung genannt:

- *Transparenz:* Industriefirmen besitzen ein großes Interesse daran, dass die Algorithmen auf den firmeninternen Rechnern bekannt sind. So sind eventuelle Modifikationen stets möglich, etwa im Bereich der Mensch-Maschinen-Schnittstelle. Dieser Punkt steht jedoch immer im Konflikt mit den kommerziellen Interessen der Vertreiber von Regelkreisdiagnostik.
- *Direkte Implementierung im Prozessleitsystem:* Im industriellen Umfeld bestehen häufig Probleme im Bereich der Software Schnittstellen (z.B. OPC). Aus diesem Grund und aus Gründen der Transparenz ist eine Einbettung in das bestehende Prozessleitsystem ein häufiger industrieller Wunsch.
- *Online-Einsatz:* Die Online-Verfügbarkeit der Diagnostik ergibt sich zwar bereits indirekt aus den genannten Anforderungen, sei hier jedoch noch einmal ausdrücklich erwähnt.

Abbildung 1.1 zeigt eine mögliche Einbindung des anlageweiten Diagnosesystems in bestehende Bushierarchien. Die Diagnostik arbeitet dabei online mit den aktuellen Betriebsdaten auf einem separaten Rechner, der mit diverser mathematischer Statistiksoftware ausgestattet ist. Die Kommunikation mit den Servern des Prozessleitsystems funktioniert mittels *OPC*. Da diese, wie beschrieben, jedoch häufig nicht gewährleistet ist, soll das Diagnosesystem auch direkt auf den Servern des Prozessleitsystems implementierbar sein. Dies zieht allerdings eine erhebliche Einschränkung der zur Verfügung stehenden Funktionalität mit sich.

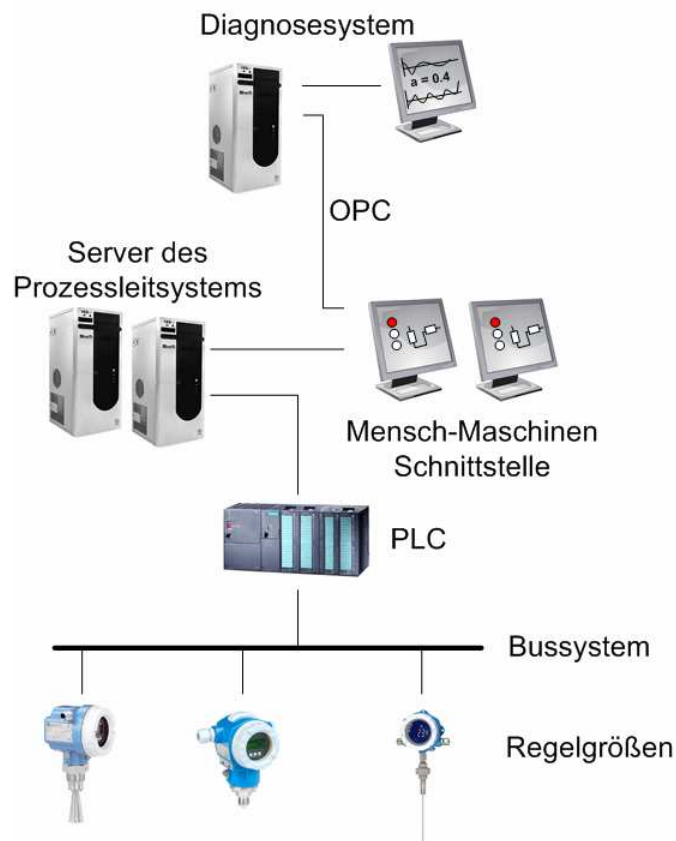


Abbildung 1.1: Optimale Einbindung des anlagenweiten Diagnosesystems in eine bestehende Anlage

1.3 Motivation und Zielsetzung

Bereits geringe Verbesserungen der Regelungsqualität führen in der Industrie oft dazu, dass Regelungen näher am Optimum laufen können oder aber die Produktqualität, z.B. durch weniger betriebliche Schwankungen, in geringem Maße verbessert wird. Diese geringen Verbesserungen bieten in Kombination mit hohen industriellen Durchsätzen ein enormes Gewinnpotential bei geringen Investitionskosten. Somit ist es offensichtlich, dass es ein großes Interesse vonseiten der Großindustrie an der anlagenweiten Regelkreisdagnostik gibt, jedoch nur unter Wahrung der in Abschnitt 1.2 genannten Anforderungen und Wünschen. Die steigende Rechnerleistung und der enorme Umfang mathematischer Software haben in den letzten Jahren dazu geführt, dass es auf dem Markt eine Vielzahl an kommerziell vertriebenen Programmen zur automatischen Regelkreisd Diagnose gibt. Zu den bekanntesten zählen Loop ScoutTM von der Firma Honeywell, Process doctorTM der Firma Matrikon und Loop Optimizer SuiteTM von der Firma ABB. Zum heutigen Zeitpunkt nutzt aber kein Programm die Möglichkeiten des anlagenweiten Ansatzes vollständig aus und außerdem verstoßen zusätzlich all diese Programme zumindest gegen eine der genannten industriellen Anforderungen. Sind zum Beispiel umfassende Analysen gewünscht, so ist der Nutzer stets vor die Wahl gesetzt, entweder ein Programm zu nutzen, welches eine vorhandene OPC Verbindung voraussetzt oder aber ein nicht automatisierbares Offline-Programm zu verwenden, was häufig lediglich von geschultem Personal anwend- und interpretierbar ist. Lediglich stark vereinfachte Analysen sind derzeit nach [12] PLS-intern möglich. Aus Gründen der Geheimhaltung ist der Programmcode keiner

kommerziellen Software bekannt, was bedeutet, dass kein kommerzielles Programm vollständig transparent ist. Ein weiterer Grund, der gegen die kommerziellen Programme spricht, sind die hohen Anschaffungskosten, die oft mit laufenden Lizenzgebühren verbunden sind. Auch wenn in der Literatur keine exakten Zahlen auffindbar sind, belaufen sich die Kosten für Matrikons Process Doctor Schätzungen nach auf mehr als 35.000\$ in der Anschaffung. Die Zielsetzung der vorliegenden Arbeit ist es somit, ein Gesamtkonzept zur anlagenweiten Regelkreisdiagnostik zu entwickeln, welches den genannten industriellen Anforderungen entspricht und für jeden Anlagenfahrer verständliche Ergebnisse liefert ohne enorme Investitionskosten mit sich zu bringen.

Die Einbindung eines anlagenweiten Ansatzes liefert zusätzlich den Vorteil, dass es in das bestehende Alarmmanagement einer Firma integriert werden kann. Der Grund hierfür liegt in der Tatsache, dass anlagenweite Fehler, welche durch einen Verursacher bedingt sind, in der Regel zu Alarmen in mehreren Sektionen führen. Durch die Identifikation des Verursachers ist es möglich, die Fülle der Alarme zu reduzieren und lediglich den Alarm im verursachenden Regelkreis zu generieren. Die aus der EEMUA 191 [14] abgeleitete Namur Empfehlung NA 102 [45] legt beispielsweise fest, dass maximal 10 Alarme in 10 Minuten im Störfall zulässig sind. Es ist offensichtlich, dass die Einhaltung dieser Empfehlung im Falle eines anlagenweiten Fehlers unmöglich ist. Eine Reduzierung der unnötigen Alarme führt nach [20] ebenfalls zu einer erhöhten Anlagenverfügbarkeit wodurch erneut der Wirkungsgrad der Anlage heraufgesetzt wird. Außerdem trägt es der industriellen Anforderung an Priorisierung Rechnung.

Um Fehlanalysen zu vermeiden, liegt ein Hauptaugenmerk der vorliegenden Arbeit ebenfalls auf dem Testen und Validieren des entwickelten Gesamtkonzepts an simulierten und industriellen Benchmarks. Das ausdrückliche Online-Reglertuning soll nicht Teil der Arbeit sein, da es auf diesem Gebiet bereits zahlreiche Veröffentlichungen gibt, wie z.B. [47] und [67]. Das Konzept soll lediglich im Schritt der Diagnose aggressive und langsame Regler erkennen. Die anschließende Regleroptimierung ist dann den genannten Quellen zu entnehmen.

1.4 Organisation der Dissertation

Im zweiten Kapitel werden bekannte und neu entwickelte Verfahren zur Identifizierung anlagenweiter Fehlerverursacher gezeigt. Diese unterscheiden sich in ihrer Laufzeit und können somit je nach Anwendungsgebiet ausgewählt werden.

In Kapitel 3 werden Ansätze zur Fehlererkennung genannt und Kapitel 4 umfasst die eigentliche Fehlerdiagnose, welche sowohl bekannte Verfahren als auch neu entwickelte Verfahren enthält. Aufgrund der Komplexität wird die Diagnose von Ventilen mit Haft- und Gleitreibung gesondert in Abschnitt 4.3 beschrieben, wohingegen alle anderen Regelungsszenarien in Abschnitt 4.2 vorgestellt werden.

Um eine möglichst hohe Zuverlässigkeit zu gewährleisten, wird das Gesamtkonzept der Qualitätskontrolle und Diagnose in Kapitel 5 an einigen simulierten und industriellen Benchmark-Prozessen getestet. Im letzten Kapitel 6 wird dann eine Zusammenfassung der Ergebnisse gegeben. Zur Vereinfachung ist die Kapitelorganisation in Abbildung 1.2 gezeigt.

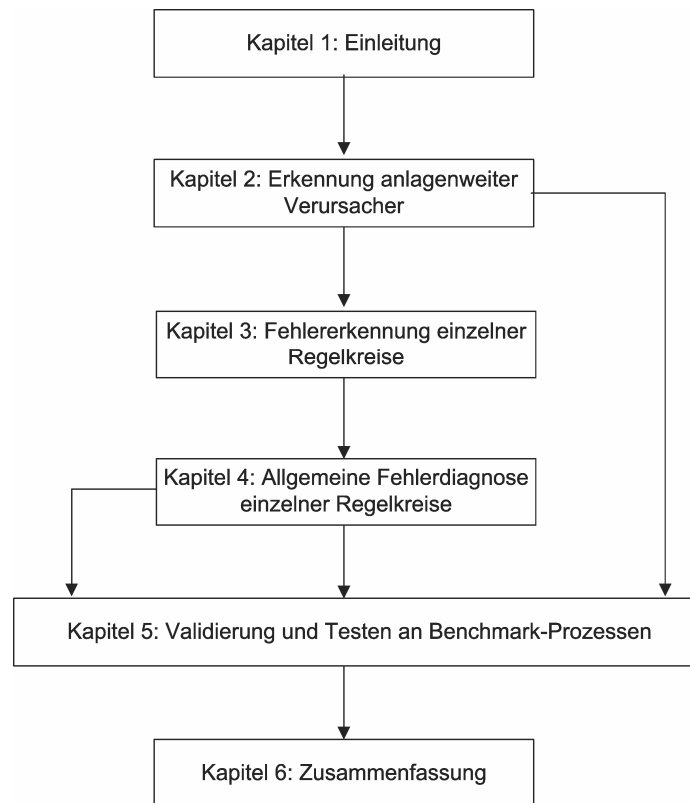


Abbildung 1.2: Organisation der Dissertation

Kapitel 2

Erkennung anlagenweiter Fehlerverursacher

Heutzutage wird die Prozedur der Fehlererkennung und der anschließenden Fehlerdiagnose häufig um einen komplexen und rechenintensiven anlagenweiten Ansatz ergänzt. Zunehmende Rechenleistung moderner Computer und die gesteigerte Leistungsfähigkeit mathematischer Software spielen hierbei die wahrscheinlich wichtigste Rolle. Der Kernpunkt eines solchen anlagenweiten Ansatzes ist es, jene Regelkreise, welchen einen Fehler lediglich durch die Beeinflussung eines anderen Regelkreises zeigen, von dem verursachenden Regelkreis zu trennen, bevor die eigentliche Fehlererkennung und Fehlerdiagnose durchgeführt wird. Der wesentliche Vorteil eines anlagenweiten Ansatzes liegt also darin, dass nicht alle Regelkreise einzeln diagnostiziert werden müssen, sondern lediglich der verursachende Regelkreis.

Zum besseren Verständnis der Vorgehensweise, stelle man sich eine Anlage, bestehend aus mehreren Regelkreisen, vor, die im Kreislaufbetrieb gefahren wird. Eine einzelne aggressive Regelung wird im Allgemeinen dazu führen, dass die zugehörige Regelgröße zu schwingen beginnt. Der Produktstrom verbreitet diese Schwingung anschließend durch die gesamte Anlage. Der klassische Ansatz, bei dem jeder Regelkreis einzeln untersucht wird, erkennt im ersten Schritt der Fehlererkennung in jedem Regelkreis zumindest die Oszillation. Im schlimmsten Fall und abhängig von der Untersuchungsart wird die anschließende Fehlerdiagnose im zweiten Schritt auch in jedem Regelkreis das Problem „Falsche Reglereinstellung“ diagnostizieren.

An diesem einfachen Beispiel lässt sich der wissenschaftliche Nutzen des anlagenweiten Ansatzes erkennen, der darin besteht, dass die Ergebnisse der Diagnose in aller Regel genauer und verlässlicher werden. Der wirtschaftliche Nutzen dieses Ansatzes lässt sich in [29] finden, da die durchschnittliche Dauer für die ausführliche Diagnose eines Regelkreises zwischen zwei und vier Stunden beträgt und eine durchschnittliche chemische Anlage z.B. aus 2000 Regelkreisen besteht.

Auch wenn der anlagenweite Ansatz erst in den letzten Jahren an Bedeutung gewonnen hat, gibt es jedoch schon mehrere Herangehensweisen. Die bekanntesten nutzen dabei entweder die Fehlerfortpflanzungsanalyse [2], frequenzbasierte Verfahren [68], [66], [72] oder Verfahren basierend auf Qualitätsindizes (z.B. Nichtlinearitätsindex [6]). In diesem Kapitel werden zwei selbst entwickelte Ansätze zur Identifikation von Verursachern anlagenweiter Fehler gezeigt. Der erste Ansatz, welcher im Folgenden gezeigt wird, gehört zur Fehlerfortpflanzungsanalyse [74]. Er ist, bedingt durch seinen hohen Berechnungsaufwand, hauptsächlich für den Offline-Einsatz konzipiert und identifiziert die Versucher von oszillierenden, stochastischen und arbiträren Fehlern anhand der Beeinflussungsrichtung. Die Grundlage für diesen Ansatz

ist, dass der aktive Verursacher eines anlagenweiten Fehlers die restlichen Regelkreise mehr beeinflusst, als diese lediglich passiven Regelkreise den Verursacher beeinflussen. Um gegebenenfalls doch einen Online-Einsatz zu ermöglichen, werden für diesen Ansatz zwei einzelne Teilansätze vorgestellt, die sich in Ihrem Rechenbedarf noch einmal signifikant unterscheiden. Zum einen basierend auf der Kreuzkorrelationsanalyse für lediglich lineare Prozesse und zum anderen basierend auf einem entwickelten rechenintensiven Verfahren, welches die k-nächste-Nachbarn-Ersetzung nutzt für nichtlineare Prozesse und MISO-Prozesse (*Multi-Input Single-Output*, Multiple Eingänge - Einzelner Ausgang). Der zweite Ansatz gehört zu den frequenzbasierten Verfahren und ist hauptsächlich für den Online-Einsatz im Prozessleitsystem entwickelt worden [72]. Es identifiziert die Verursacher von sinusförmig oszillierenden Fehlern basierend auf der Analyse der übertragenen Leistung im Frequenzband des anlagenweiten Fehlers.

2.1 Problemstellung

Die Erkennung von anlagenweiten Fehlern wird exemplarisch für die generalisierte und abstrahierte Anlage in Abbildung 2.1 gezeigt, welche aus M Teilprozessen (z.B. Regelkreisen) S_1, S_2, \dots, S_M besteht. Alle Teilprozesse sind dabei entweder direkt oder indirekt miteinander verbunden. Zusätzlich passieren die Ausgänge aller Prozesse definierte und zeitinvariante Totzeiten $d_{0,1}, d_{1,2}, \dots, d_{M,1}$ auf ihrem Verbindungsweg. Bei S_0 handelt es sich um einen externen Eingang, der in der Praxis oft nicht messbar ist. Es ist offensichtlich, dass diese Anlage eine determinierte Hierarchie besitzt. Der erste Teilprozess S_1 wirkt direkt auf S_2 , S_2 auf S_3 oder allgemein S_i wirkt direkt auf S_{i+1} . Neben diesen direkten Wirkungswegen wirkt S_i allerdings auch indirekt auf S_{i+2} bis S_M und durch den Kreislauf auch auf S_j bis S_{i-1} . Die potentielle Verbreitung eines anlagenweiten Fehlers ist somit gegeben, da jeder Teilprozess S_i jeden anderen Teilprozess S_j beeinflusst.

Die folgenden Vorgehensweisen (Offline- und Online-Ansatz) sind lediglich konzipiert für einzelne anlagenweite Fehler. Es soll nicht Teil dieser Arbeit sein, die Fragestellung zu klären, wie die beiden Ansätze im Falle von mehreren simultanen anlagenweiten Fehlern arbeiten. Ein Beispiel hierfür ist eine aggressive Regelung in einem Regelkreis und eine begrenzte Regelung in einem anderen Regelkreis.

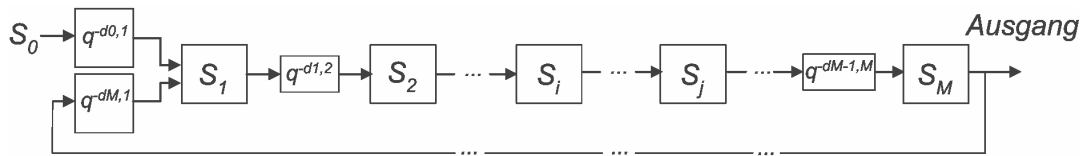


Abbildung 2.1: Abstrahierte Anlage mit M Teilprozessen und Kreislaufführung [74]

Bemerkung: Die Erkennung des Verursachers eines anlagenweiten Fehlers ist nicht auf die Anlagenbeschreibung in Abbildung 2.1 beschränkt, allerdings zeigt die Erfahrung, dass viele chemische Anlagen oder Teilanlagen mit dieser Beschreibung gut wiedergegeben werden können. Typische Gründe für einen solchen Kreislaufprozess können dabei Wärmekopplung oder stoffliche Kreisläufe sein. Für den Fall, dass eine zu untersuchende Anlage von der generalisierten Anlage abweicht, muss vor der eigentlichen Identifikation des Verursachers eine Klassifikation in zwei Gruppen durchgeführt werden. Gruppe 1 besteht dabei aus allen Teilprozessen, welche den

betrachteten anlagenweiten Fehler aufweisen (Fehlererkennung, s. Kapitel 3) und Gruppe 2 aus allen restlichen. Die Erkennung des Verursachers muss anschließend lediglich für alle Teilprozesse aus Gruppe 1 durchgeführt sein. Dieses Vorgehen ist bei der generalisierten Anlage nicht anzuwenden, da anzunehmen ist, dass es hier keine Gruppe 2 gibt.

2.2 Erkennung im Offline-Betrieb

Es wurde gezeigt, dass für die verallgemeinerte Anlage gilt, dass jeder Teilprozess S_i jeden anderen Teilprozess S_j beeinflusst. Um die Bedeutung dessen für die Erkennung eines Verursachers zu klären, seien folgende zwei Fälle anzunehmen: Im ersten Fall arbeitet die gesamte Anlage normal ohne anlagenweiten Fehler. Die maßgeblich beeinflussende Größe ist lediglich der Anlageneingang S_0 , welcher nicht messbar sei, jedoch direkt verbunden mit S_1 ist. Deshalb ist es offensichtlich, dass S_i den nachfolgenden Teilprozess S_j mehr beeinflusst als S_j den vorangehenden Teilprozess S_i beeinflusst. Für den zweiten Fall sei anzunehmen, dass keine signifikanten Störungen durch S_0 in die Anlage gebracht werden und dass es zusätzlich einen anlagenweiten Fehler gibt, hervorgerufen durch S_j . In diesem Fall wird die hauptsächliche Beeinflussungsrichtung umgekehrt sein, da der beeinflussende Teilprozess in diesem Fall hauptsächlich S_j ist. Das bedeutet, dass S_j jeden anderen Teilprozess mehr beeinflusst als er selbst von anderen Teilprozessen beeinflusst wird. Der Verursacher eines anlagenweiten Fehlers lässt sich somit dadurch erkennen, dass die Messung dieses Teilprozesses als einzige alle anderen Messungen der restlichen Teilprozesse beeinflusst. Um dies zu visualisieren, wird die *Beeinflussungsmatrix* \mathbf{Z} eingeführt. Für die M Teilprozesse der verallgemeinerten Anlage ergibt sich eine Matrix \mathbf{Z} mit den Dimensionen $M \times M$. Jede Zeile und jede Spalte steht für die Messung eines jeden Teilprozesses und der i,j -te Eintrag gibt die Nummer (i oder j) oder den Namen des Teilprozesses an, der den anderen Teilprozess signifikant mehr beeinflusst. Diese Definition führt dazu, dass \mathbf{Z} eine symmetrische Matrix mit keinen Einträgen auf der Hauptdiagonalen ist. Um die generelle Aufstellung der Matrix darzustellen, seien noch einmal die beiden vorherigen Fälle aufgegriffen. Im ersten Fall, wird die gesamte Anlage hauptsächlich durch externe Störungen beeinflusst. Dies führt dazu, dass sich die Beeinflussungsmatrix \mathbf{Z}_1 ergibt gemäß

$$\mathbf{Z}_1 = \begin{matrix} & \begin{matrix} S_1 & S_2 & S_3 & \dots & S_i & \dots & S_j & \dots & S_M \end{matrix} \\ \begin{matrix} S_1 \\ S_2 \\ S_3 \\ \vdots \\ S_i \\ \vdots \\ S_j \\ \vdots \\ S_M \end{matrix} & \begin{bmatrix} - & 1 & 1 & \dots & 1 & \ddots & 1 & \dots & 1 \\ 1 & - & 2 & \dots & 2 & \ddots & 2 & \dots & 2 \\ 1 & 2 & - & \dots & 3 & \ddots & 3 & \dots & 3 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 1 & 2 & 3 & \dots & - & \ddots & i & \dots & i \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 1 & 2 & 3 & \dots & i & \ddots & - & \dots & j \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 1 & 2 & 3 & \dots & i & \ddots & j & \dots & - \end{bmatrix} \end{matrix}, \quad (2.1)$$

welche in der ersten Zeile und der ersten Spalte nur aus dem Eintrag „1“ besteht.

Für den zweiten Fall mit dem Verursacher eines anlagenweiten Fehlers in S_j ergibt sich die Matrix \mathbf{Z}_2 gemäß

$$\mathbf{Z}_2 = \begin{matrix} & S_1 & S_2 & S_3 & \dots & S_i & \dots & S_j & \dots & S_M \\ \begin{matrix} S_1 \\ S_2 \\ S_3 \\ \vdots \\ S_i \\ \vdots \\ S_j \\ \vdots \\ S_M \end{matrix} & \begin{bmatrix} - & 1 & 1 & \dots & 1 & \ddots & j & \dots & M \\ 1 & - & 2 & \dots & 2 & \ddots & j & \dots & M \\ 1 & 2 & - & \dots & 3 & \ddots & j & \dots & M \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 1 & 2 & 3 & \dots & - & \ddots & j & \dots & M \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ j & j & j & \dots & j & \ddots & - & \dots & j \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ M & M & M & \dots & M & \ddots & j & \dots & - \end{bmatrix} \end{matrix}, \quad (2.2)$$

wobei die j -te Zeile und Spalte nur den Eintrag „ j “ enthält.

Wenn es somit einen Ansatz zur Klärung der Beeinflussungsrichtung gibt, ist es generell möglich, den Verursacher zu identifizieren, indem nach j -ten Zeilen und j -ten Spalten gesucht wird, die nur den Eintrag j enthalten. Vergleicht man den Fall der externen Beeinflussung in (2.1) mit dem Fall des internen Verursachers in (2.2), so wird eine Einschränkung dieser Herangehensweise erkenntlich. Im Fall der externen Beeinflussung besteht die erste Zeile und erste Spalte lediglich aus dem Eintrag „1“. Es könnte bei normalem Betrieb somit fehlerhaft geschlussfolgert werden, dass es einen anlagenweiten Fehler gibt mit dem Verursacher im Teilprozess S_1 . Der Grund hierfür ist, dass dieser Teilprozess direkt durch die eigentliche Störung beeinflusst wird und somit folglich alle anderen Teilprozesse mehr beeinflusst, als er selbst beeinflusst wird. Zur Vermeidung dieses Problems seien zwei Ansätze genannt, die sich in ihrer Komplexität unterscheiden.

Einfache Herangehensweise: Unter Inkaufnahme einer gewissen Unsicherheit besteht diese Herangehensweise darin, anzunehmen, dass anlagenweite Fehler niemals im ersten Teilprozess S_1 entstehen. Dies hat zur Folge, dass Beeinflussungsmatrizen wie \mathbf{Z}_1 den normalen Betrieb darstellen. Jede andere Matrix mit einem identifizierten Verursacher in S_j ($j \neq 1$) führt zu einer anschließenden ausführlichen Diagnose des Teilprozesses S_j um den eigentlichen Grund des Fehlers zu klären. Es ist offensichtlich, dass es sich bei dieser Herangehensweise um eine sehr starke Vereinfachung mit dem Risiko einer Fehlinterpretation handelt. Es muss allerdings gesagt werden, dass dieser Ansatz im industriellen Umfeld ebenfalls seine Berechtigung hat, da die Wahrscheinlichkeit für einen anlagenweiten Fehler in lediglich einem aus M Teilprozessen oft sehr gering ist.

Komplexe Herangehensweise: Um wissenschaftlichen Ansprüchen zu genügen, wurde ein weiterer komplexerer Ansatz entwickelt, dessen Ablauf in Abbildung 2.2 zu sehen ist. Im ersten Schritt muss die Beeinflussungsmatrix berechnet werden. Gibt es einen identifizierten Verursacher, so hängt das weitere Vorgehen von dessen Lage ab. Handelt es sich bei dem identifizierten Verursacher um S_1 und ist S_0 nicht bekannt, so muss eine ausführliche Fehlerdiagnose durchgeführt werden. Liefert

diese kein Ergebnis, so läuft die Anlage vermutlich normal und wird lediglich durch äußere Störungen beeinflusst.

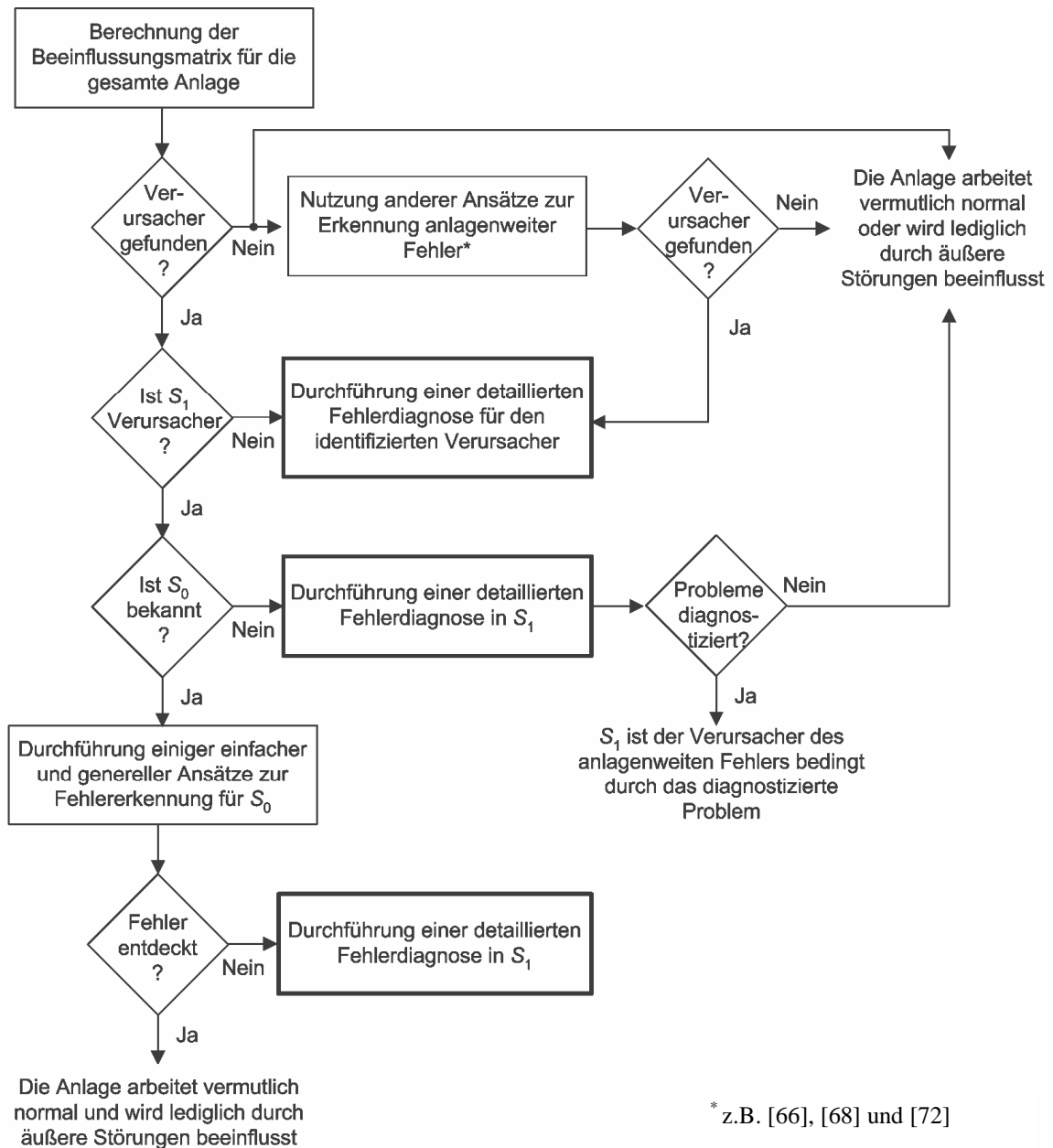


Abbildung 2.2: Prozedur zur Identifikation von anlagenweiten Fehlerverursachern [74]

Die Klärung der hauptsächlichen Beeinflussungsrichtung ist, sofern es keinerlei Einschränkungen gibt (z.B. linear, messbare Totzeit) eine schwierige mathematische Fragestellung. So nutzt [2] etwa die für die Informationstechnologie entwickelte Shannon-Entropy nach [53] zwischen zwei Signalen u und y um zu bestimmen, welches Signal das andere mehr beeinflusst. Die Shannon-Entropy h_1 für zwei Signale ist definiert gemäß

$$h_1 = - \sum_{u,y} p(y,u) \cdot \log(p(y,u)) . \quad (2.3)$$

Sind beide Signale stochastisch unabhängig, lässt sich (2.3) umformulieren als

$$h_2 = - \sum_{u,y} p(y,u) \cdot \log(p(y) \cdot p(u)), \quad (2.4)$$

wobei $\sum_{u,y}(\cdot)$ angibt, dass die Berechnung der Shannon-Entropien h_1 und h_2 durch

Summation aller Histogrammklassen (für $p(u)$, $p(y)$ und $p(u,y)$) erfolgt, aus denen die empirischen Wahrscheinlichkeiten abgelesen werden können. In der Literatur sind allerdings keine Regeln zur Wahl dieser Klassen zu finden. Basierend auf (2.3) und (2.4) kann die Differenz der beiden Entropien nach [52] als Maß für die wechselseitige Beziehung genutzt werden. Diese ergibt sich somit nach

$$h_2 - h_1 = \sum_{u,y} p(y,u) \cdot \log \frac{p(y,u)}{p(y) \cdot p(u)}. \quad (2.5)$$

Ist die Differenz beider Entropien Null, so gilt offensichtlich, dass beide Signale stochastisch unabhängig sind. Bedingt durch die Tatsache, dass $h_2 - h_1$ jedoch symmetrisch ist ($p(y,u) = p(u,y) \wedge p(y) \cdot p(u) = p(u) \cdot p(y)$), gilt, dass die einfache Differenz nicht genutzt werden kann, um die Richtung des Einflusses zu klären.

Zur Berücksichtigung der Richtung, wird bei der so genannten Transferentropie die Information der nächsten Beobachtung y_{k+1} (oder u_{k+1}) mit berücksichtigt unter der Voraussetzung, dass y_k und u_k aufgetreten sind. Da dieser Ausdruck entweder die Wahrscheinlichkeit für das Auftreten von y_{k+1} oder u_{k+1} berücksichtigt, ist dieser nicht mehr symmetrisch, so dass die zwei Transferentropien aufgestellt werden können nach

$$t_{u \rightarrow y} = \sum p(y_{k+1}, y_k, u_k) \cdot \log \frac{p(y_{k+1} | y_k, u_k)}{p(y_{k+1} | y_k)} \quad \text{und} \quad (2.6)$$

$$t_{y \rightarrow u} = \sum p(u_{k+1}, y_k, u_k) \cdot \log \frac{p(u_{k+1} | y_k, u_k)}{p(u_{k+1} | u_k)}. \quad (2.7)$$

Die Transferentropien in (2.6) und (2.7) sind somit ein Maß für die stochastische Abhängigkeit und für die Beeinflussungsrichtung zwischen u_k , y_k und y_{k+1} , bzw. u_k , y_k und u_{k+1} . Es gilt die Beziehung, dass sofern

$t_{u \rightarrow y} > t_{y \rightarrow u}$ gilt, u die Zeitreihe y mehr beeinflusst und analog sofern

$t_{u \rightarrow y} < t_{y \rightarrow u}$ gilt, y die Zeitreihe u mehr beeinflusst.

Der Nachteil dieses Ansatzes ist allerdings, dass ein unmittelbarer Zusammenhang zwischen u und y existieren muss, da lediglich die bedingte Wahrscheinlichkeit von y_{k+1} , bzw. u_{k+1} bestimmt wird. Eine gewisse Totzeit zwischen beiden Signalen würde das Ergebnis somit verfälschen. Als weiterer Nachteil kann die Tatsache angesehen werden, dass lediglich das Auftreten von y_k und u_k vorausgesetzt wird. Besser (z.B. bei Prozessen hoher Ordnung) ist die Berücksichtigung der l letzten u Werte und der m

letzten y Werte. [2] ändert somit die Transferentropien abhängig von einer diskreten Verschiebung n_d und unter Berücksichtigung des genannten Horizonts ab gemäß

$$t_{u \rightarrow y}(n_d) = \sum p(y_{k+nd}, \mathbf{y}_k^m, \mathbf{u}_k^l) \cdot \log \frac{p(y_{k+nd} | \mathbf{y}_k^m, \mathbf{u}_k^l)}{p(y_{k+nd} | \mathbf{y}_k^m)}, n_d \geq 1 \text{ und} \quad (2.8)$$

$$t_{y \rightarrow u}(n_d) = \sum p(u_{k+nd}, \mathbf{y}_k^m, \mathbf{u}_k^l) \cdot \log \frac{p(u_{k+nd} | \mathbf{y}_k^m, \mathbf{u}_k^l)}{p(u_{k+nd} | \mathbf{u}_k^l)}, n_d \geq 1. \quad (2.9)$$

Diese Modifikation hat allerdings zur Folge, dass der Berechnungsaufwand drastisch zunimmt. Um dies zu verdeutlichen, sei das folgende Beispiel gezeigt [70]: Es ist anzunehmen, dass y_k und u_k binäre Signale seien, wodurch das Histogramm beider Größen aus zwei Klassen (0 und 1) besteht. Somit wird das Histogramm zur Berechnung von $p(y_{k+1}, y_k, u_k)$ dreidimensional und enthält 2^3 Histogrammklassen.

Wird nun y_k durch \mathbf{y}_k^m und u_k durch \mathbf{u}_k^l ersetzt mit $m = l = 3$ folgt daraus, dass es für diesen binären Fall 2^7 Histogrammklassen geben muss. Für Prozesse höherer Ordnung kann die Wahl $m = l = 3$ sogar zu gering sein, was dazu führt, dass schnell eine Anzahl an Histogrammklassen aufgestellt wird, die unmöglich alle realisiert werden. Dies führt dazu, dass gewisse Klassen keine relativen Häufigkeiten zugewiesen bekommen. Noch umfangreicher wird die Berechnung für quasi stetige Signale; hierbei ist eine optimale Wahl der Klassenattribute essentiell. [2] gibt jedoch keinerlei Anhaltspunkte über eine geeignete Wahl der Größen l, m oder der Klassenattribute.

2.2.1 Erkennung im Falle von ausschließlich linearen Teilprozessen

Es wurde abstrahiert gezeigt, wie die Beeinflussungsmatrix zur Identifikation von Verursachern anlagenweiter Fehler genutzt werden kann und wie diese mittels Transferentropie allgemein bestimmt werden kann. Bedingt durch die Tatsache, dass zur Berechnung der Transferentropie viele Parameter (Klassenattribute, Vektorgröße, ...) variabel sind und nur bei optimaler Wahl das richtige Ergebnis liefern, bietet sich dieses Verfahren zur Klärung der Beeinflussungsrichtung nicht an, da es den Anforderungen nach Automatisierbarkeit aus Abschnitt 1.2 widerspricht. Um diesen Anforderungen zu genügen, wird in diesem Abschnitt ein Verfahren entwickelt, welches die Einschränkungen annimmt, dass alle Teilprozesse linear sind und dass die Totzeiten zwischen allen Teilprozessen größer sind als die Abtastzeit.

Die Hauptidee des neuen Verfahrens ist, dass die Fragestellung der Beeinflussungsrichtung zwischen S_i und S_j im Falle einer messbaren Totzeit auch durch die Fragestellung „Welches Signal (S_i oder S_j) entstand zuerst“ ersetzt werden kann. Die Gleichheit beider Fragestellungen wird in diesem Abschnitt gezeigt. Eine Klärung dieser vereinfachten Frage lässt sich durch eine Totzeitschätzung zwischen S_i und S_j beantworten. Unter der Voraussetzung, dass alle Teilprozesse linear sind, ist die Totzeit durch Berechnung der Kreuzkorrelationsfunktion schätzbar. Hierbei handelt es sich um das bekannteste Verfahren. Die Kreuzkorrelationsfunktion zwischen S_i und S_j lässt sich auf zwei Arten, abhängig vom angenommenen Eingang, schätzen.

Mit dem angenommenen Eingang S_i , dessen skalierte Messung s_i^* der Länge n und dem sich daraus ergebenden Ausgang S_j mit dessen skalierte Messung s_j^* berechnet

sich die Kreuzkorrelationsfunktion in Abhängigkeit einer diskreten Verschiebung n_d gemäß

$$ccf_{S_i, S_j}(n_d) = \frac{1}{n - n_d} \sum_{k=1}^{n-n_d} s_i^*(k - n_d) \cdot s_j^*(k). \quad (2.10)$$

Für den angenommenen Eingang S_j und dessen skaliertes Messung s_j^* berechnet sie sich gemäß

$$ccf_{S_j, S_i}(n_d) = \frac{1}{n - n_d} \sum_{k=1}^{n-n_d} s_j^*(k - n_d) \cdot s_i^*(k). \quad (2.11)$$

Die eigentliche Schätzung der Totzeit im Intervall $n_d \in [d_{\min}, d_{\max}]$ mit einer minimal anzunehmenden Totzeit d_{\min} und einer maximal anzunehmenden Totzeit d_{\max} geschieht nach

$$\hat{d} = \arg \max_{n_d \in [d_{\min}, d_{\max}]} |ccf_{S_i, S_j}(n_d)| \quad \vee \quad \hat{d} = \arg \max_{n_d \in [d_{\min}, d_{\max}]} |ccf_{S_j, S_i}(n_d)|. \quad (2.12)$$

Im Falle einer negativen Totzeit ($\text{sign}(\hat{d}) = -1$) gilt, dass die Annahme bezüglich des Eingangs falsch ist und der andere Teilprozess als Eingang angesehen werden kann und somit dessen Signal als erstes entstand. Im Falle einer geschätzten positiven Totzeit ist die Annahme bezüglich des Eingangs korrekt. Diese Vorgehensweise sei für den einfachen linearen und statischen Prozess ohne Kreislauf

$$s_{j,k} = -1 + 0.5s_{i,k-3} \quad (2.13)$$

gezeigt. Die Messung s_i besteht dabei aus 100 normalverteilten Zufallszahlen mit dem Mittelwert 0 und der Varianz 4. Die beiden Kreuzkorrelationsfunktionen zwischen S_i und S_j sind in Abbildung 2.3 zu sehen. In a) ist der Funktionsverlauf für den angenommenen Eingang S_i und in b) für den angenommenen Eingang S_j gezeigt.

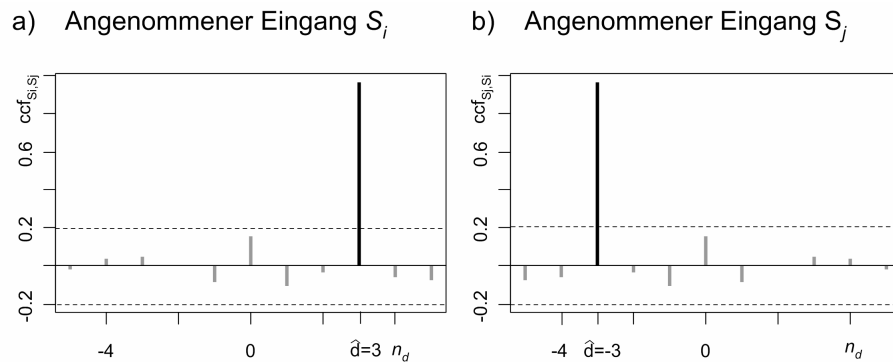


Abbildung 2.3: Kreuzkorrelationsfunktion zwischen S_i und S_j mit a) dem angenommenen Eingang S_i und b) dem angenommenen Eingang S_j (gepunktete Linie: Signifikanzniveau) [74]

An den beiden Verläufen ist zu erkennen, dass es sich bei S_i um den tatsächlichen Eingang handelt und S_i somit S_j beeinflusst. Ferner kann aus den Verläufen auch

abgelesen werden, dass S_i in keiner Weise von S_j beeinflusst wird. Dies ist zum Beispiel für Abbildung 2.3 a) daran zu erkennen, dass es für den Bereich $n_d < 0$ keine signifikanten Peaks gibt.

Im Falle eines Kreislaufprozesses mit der Umlaufzeit T_{circ} sei $s_{j,k}$ wieder durch $s_{i,k-3}$ beeinflusst, aber gleichzeitig wird $s_{i,k}$ auch beeinflusst von $s_{j,k-(T_{circ}-3)}$. Es ist also offensichtlich, dass die Kreuzkorrelationsfunktion genutzt werden kann, um zu klären, welche Einflüsse stärker sind, denn die Richtung des größeren Einflusses wird, unter der Voraussetzung linearer Teilprozesse, das Maximum im Verlauf der Kreuzkorrelationsfunktion liefern. Dies ist in Abbildung 2.4 verallgemeinert dargestellt mit dem angenommenen Eingang S_i für a) den Fall, dass S_i den Teilprozess S_j mehr beeinflusst und in b) für den Fall, dass S_j den Teilprozess S_i mehr beeinflusst.

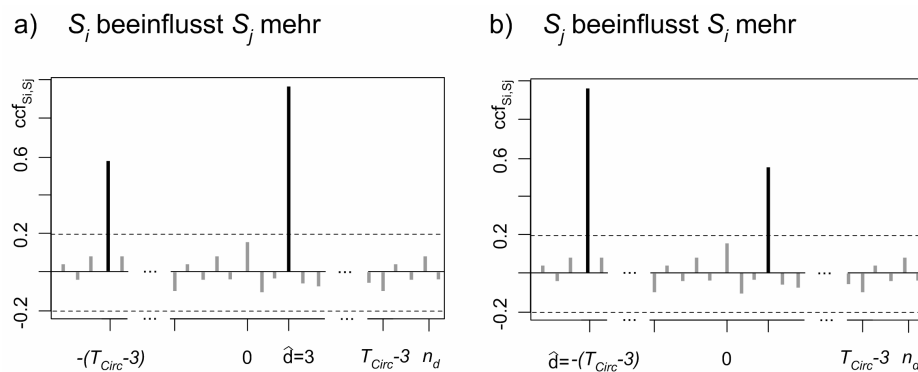


Abbildung 2.4: Kreuzkorrelationsfunktion zwischen S_i und S_j und dem angenommenen Eingang S_i (gepunktete Linie: Signifikanzniveau) [74]

Die daraus gewonnenen Erkenntnisse für die Berechnung der Beeinflussungsmatrix werden anhand eines simulierten Beispiels gezeigt.

Simulationsbeispiel 2.1: Es sei die Anlage aus Abbildung 2.5 mit $M = 3$ gegeben. Jeder Teilprozess sei linear und die Totzeiten sind mit $d_{0,1} = 1$, $d_{1,2} = 2$, $d_{2,3} = 5$, $d_{3,1} = 6$ simuliert. Für den Prozesseingang gilt $S_0 \sim N(1, 0.0001)$. Für die Störungen gilt $\varepsilon_1 \dots \varepsilon_5 \sim N(0, 0.01)$. Außerdem wird zum Ausgang des Prozesses S_2 eine Oszillation addiert, um einen Fehler zu simulieren.

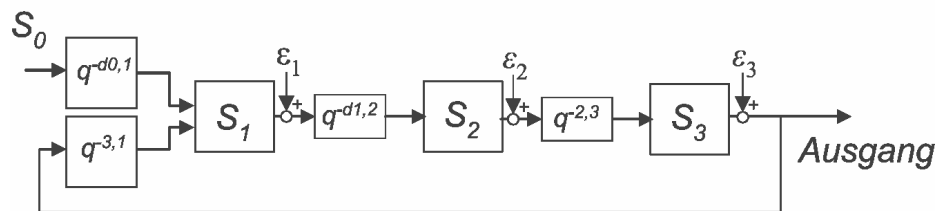
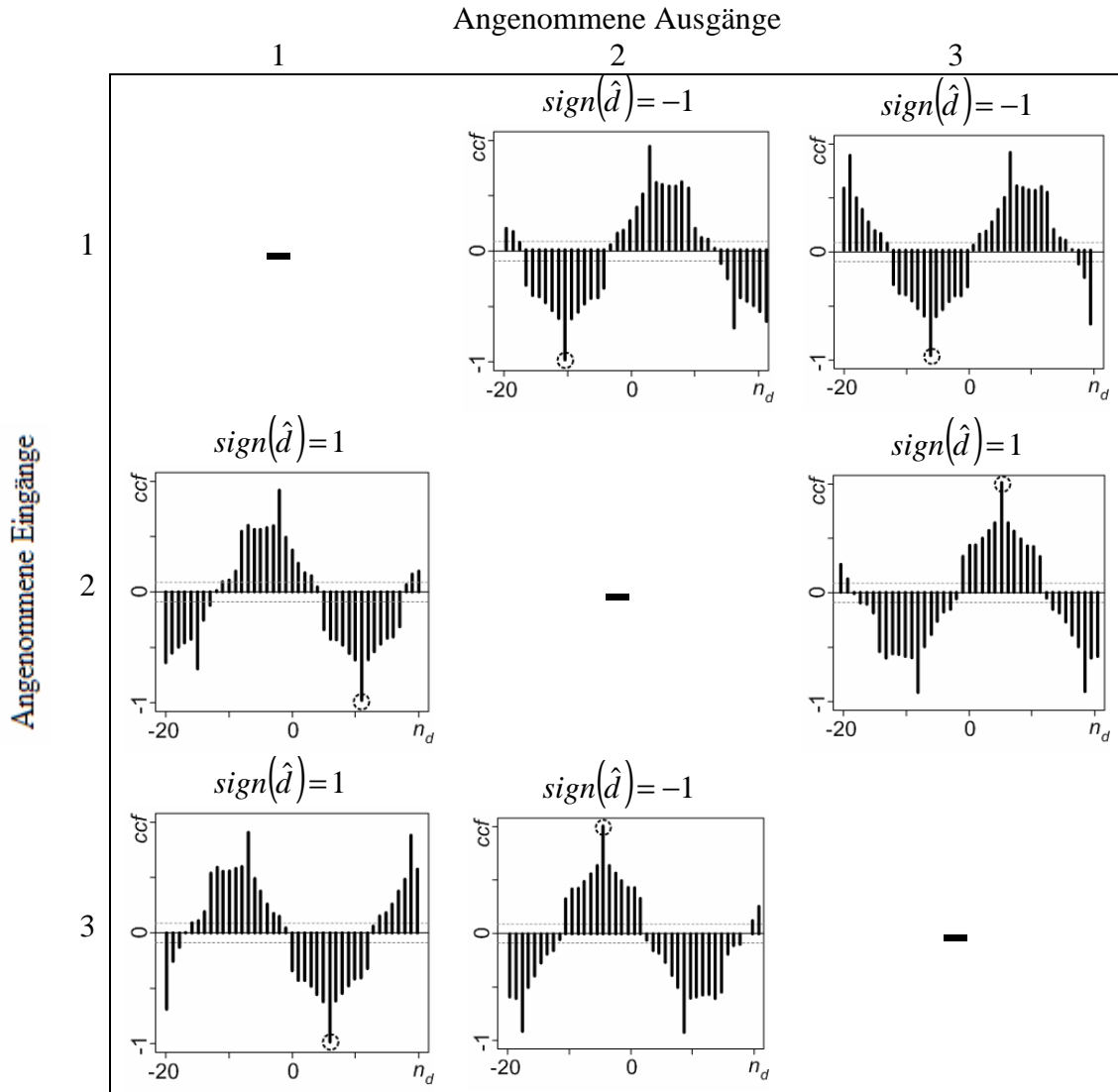


Abbildung 2.5: Anlagenmodell mit $M = 3$ linearen Teilprozessen

Tabelle 2.1: Kreuzkorrelationsfunktionen aller Permutationen



Wie bereits gezeigt wurde, müssen für alle möglichen Permutationen die Kreuzkorrelationsfunktionen bestimmt werden. Diese sind in Tabelle 2.1 zu sehen. Anhand dieser lässt sich sehr einfach das Vorzeichen der geschätzten Totzeit bestimmen. Diese sind ebenfalls dargestellt. Mit dieser Erkenntnis und den getroffenen Annahmen bezüglich Ein- und Ausgang, lässt sich aus diesen sechs Kreuzkorrelationsfunktionen die Beeinflussungsmatrix ableiten. Es reicht aus, lediglich drei Kreuzkorrelationsfunktionen zu berechnen, da die Matrix symmetrisch ist. Bei M Teilprozessen gilt dementsprechend, dass $(M^2-M)/2$ Berechnungen reichen, um \mathbf{Z} eindeutig zu bestimmen.

Die abgeleitete Beeinflussungsmatrix ergibt sich gemäß

$$\mathbf{Z} = \begin{matrix} S_1 \\ S_2 \\ S_3 \end{matrix} \begin{bmatrix} S_1 & S_2 & S_3 \\ - & 2 & 3 \\ 2 & - & 2 \\ 3 & 2 & - \end{bmatrix}. \quad (2.14)$$

Der Verursacher S_2 wurde somit korrekt identifiziert, da er sowohl S_1 als auch S_3 mehr beeinflusst, als er von diesen selbst beeinflusst wird.

2.2.2 Erweiterter Ansatz für nichtlineare Systeme und MISO-Systeme

In Abschnitt 2.2.1 wurde ein automatisierbares Verfahren zur Berechnung der Beeinflussungsmatrix \mathbf{Z} vorgestellt ([70], [73] und [74]), welches nur unter den Bedingungen anzuwenden ist, dass alle Prozesse linear sind und es sich um messbare Totzeiten handelt. Der Punkt der messbaren Totzeit wird in der Praxis oft erreicht. Dies gilt allerdings nicht für die Annahme der Linearität. Die Dynamiken vieler chemischer Prozesse sind nichtlinear und lassen sich auch nur mit sehr großen Abweichungen im Arbeitspunkt linearisieren. In diesem Abschnitt wird daher ein Verfahren vorgestellt, welches zwar auch auf einer Totzeitschätzung basiert und somit eine messbare Totzeit voraussetzt, nicht jedoch für lineare Prozesse begrenzt ist. Außerdem ist dieses Verfahren ebenfalls für dynamische Prozesse hoher Ordnung und sogar für MISO-Systeme anwendbar.

Anmerkung: Zahlreiche Publikationen (z.B. [22]) haben die Anwendbarkeit der Kreuzkorrelation auch bei nichtlinearen Prozessen bewiesen, jedoch setzt diese eine vorherige Annahme der Art der Nichtlinearität voraus. Dies widerspricht erneut den Anforderungen der kompletten Automatisierung nach 1.2.

Das Verfahren zur Totzeitschätzung sei anhand eines Systems mit einer statischen Beziehung zwischen u_k und y_{k+nd} erklärt. Ein bestimmter u_{k_0} Wert, welcher zum Zeitpunkt k_0 abgetastet wird mit dem zugehörigen Ausgang y_{k_0+nd} wird dazu führen, dass beim späteren Auftreten eines gleichen Eingangs u_{k_1} ($u_{k_0} = u_{k_1}$) zum Zeitpunkt k_1 , die Ausgänge ebenfalls gleich sind ($y_{k_0+nd} = y_{k_1+nd}$). Dies wird jedoch nur der Fall sein, wenn n_d der tatsächlichen Totzeit d zwischen u und y entspricht, sofern der Prozess nicht durch Rauschen gestört ist ($u_{k_0} = u_{k_1} \Leftrightarrow y_{k_0+nd} = y_{k_1+nd}$, wenn $n_d = d$). Dieser Effekt wird im Folgenden, basierend auf der statistischen Bedeutung, *Konsistenz* genannt.

Die Bestimmung von Konsistenz wird anhand eines nichtlinearen statischen Prozesses gezeigt. Hierfür wird angenommen, dass u_k und y_k eine Beziehung aufweisen gemäß

$$y_k = \frac{1}{2} u_{k-1}^2 + \frac{1}{2} u_{k-1} - 2. \quad (2.15)$$

Weiter sei anzunehmen, dass die folgenden Eingänge mit den sich nach (2.15) daraus ergebenden Ausgängen gemessen wurden.

$$u_k = 3, -2, 0, -3, 1, -1, 2, 4 \quad y_k = 2, 4, -1, -2, 1, -1, -2, 1$$

Zur Bestimmung der Konsistenz müssen im Folgenden die Punktmengen

$$P(n_d) = \left\{ (u_{k-n_d}, y_k) \mid k = 1, 2, \dots, n \right\} \quad (2.16)$$

der Länge n bestimmt werden unter der Bedingung, dass $n \geq k - n_d > 0$ mit $n_d \in [d_{\min}, d_{\max}]$.

Soll die unbekannte Totzeit nun geschätzt werden, muss zunächst ein n_d -Bereich festgelegt werden, in welchem nach der Totzeit gesucht werden soll. Zur Wahrung der Einfachheit sei dies für das Beispiel im Bereich von -1 bis 1, so dass sich ergibt: $n_d \in [d_{\min} = -1, d_{\max} = 1]$. Somit müssen folglich drei Punktmengen ($P(n_d = -1)$, $P(n_d = 0)$, $P(n_d = 1)$) aufgestellt werden. Für dieses Beispiel lauten sie:

$$\begin{aligned} P(-1) &= \{(u_{k+1}, y_k) \mid k = 1, 2, \dots, 8\} = (-2, 2); (0, 4); (-3, -1); (1, -2); (-1, 1); (2, -1); (4, -2) \\ P(0) &= \{(u_k, y_k) \mid k = 1, 2, \dots, 8\} = (3, 2); (-2, 4); (0, -1); (-3, -2); (1, 1); (-1, -1); (2, -2); (4, 1) \\ P(1) &= \{(u_{k-1}, y_k) \mid k = 1, 2, \dots, 8\} = (3, 4); (-2, -1); (0, -2); (-3, 1); (1, -1); (-1, -2); (2, 1) \end{aligned}$$

Die Visualisierung dieser drei Punktmengen in einem kartesischen Koordinatensystem ist in Abbildung 2.6 gezeigt.

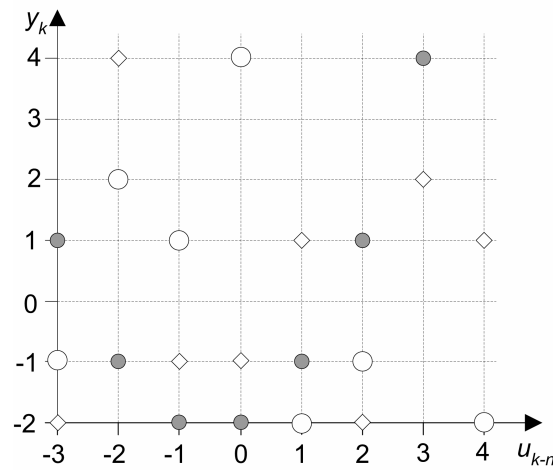


Abbildung 2.6: Visualisierung der Punktmengen $P(n_d = -1)$ (○), $P(n_d = 0)$ (◇) und $P(n_d = 1)$ (●) [74]

Die Bestimmung der Konsistenz soll nun für den Zeitpunkt $k_0 = 2$ mit dem zugehörigen Eingang $u_{k_0} = -2$ gezeigt werden. Bedingt durch die Tatsache, dass es keinen weiteren aufgezeichneten Eingang u_{k_1} gibt, bei dem gilt, dass $u_{k_0} = u_{k_1}$, ist ein direkter Vergleich der Ausgänge y_{k_0+nd} und y_{k_1+nd} nicht möglich. Es besteht allerdings die Möglichkeit, den nicht vorhandenen Eingang u_{k_1} mit seinem zugehörigen Ausgang y_{k_1+nd} basierend auf den ähnlichsten Eingängen (verglichen mit $u_{k_0} = -2$) und deren zugehörigen Ausgängen zu schätzen um somit die Gleichheit von y_{k_0+nd} und y_{k_1+nd} zu überprüfen. Die Anzahl an ähnlichsten Eingängen wird im Folgenden mit k_{imp} bezeichnet und die k_{imp} zu u_{k_0} ähnlichsten Eingänge werden mit $u_{k_0,1}, u_{k_0,2}, \dots, u_{k_0,k_{imp}}$ bezeichnet. Deren um n_d verschobenen Ausgänge dementsprechend mit $y_{k_0+nd,1}, y_{k_0+nd,2}, \dots, y_{k_0+nd,k_{imp}}$. Der Ausgang y_{k_0+nd} muss somit mit einer Schätzung \hat{y}_{k_1+nd} verglichen werden. Diese lässt sich beispielsweise durch den Mittelwert von $y_{k_0+nd,1}, y_{k_0+nd,2}, \dots, y_{k_0+nd,k_{imp}}$ bestimmen gemäß

$$\hat{y}_{k_1+nd} = \frac{1}{k_{imp}} \sum_{i=1}^{k_{imp}} y_{k_0+nd,i} = \frac{y_{k_0+nd,1}}{k_{imp}} + \dots + \frac{y_{k_0+nd,k_{imp}}}{k_{imp}}. \quad (2.17)$$

Um die Erklärung einfach zu halten, sei diese Prozedur für das Beispiel $k_0 = 2$ und $u_{k_0} = -2$ mit $k_{imp} = 2$ gezeigt. Die beiden ähnlichsten Eingänge sind $u_{k_0,1} = -3$ und $u_{k_0,2} = -1$ (bzw. umgekehrt). Die zugehörigen Ausgänge hängen von der Verschiebung ab und lauten hier:

$$\begin{aligned} P(n_d = -1): y_{k_0+nd,1} &= y_{4-1} = -1 \wedge y_{k_0+nd,2} = y_{6-1} = 1 \\ P(n_d = 0): y_{k_0+nd,1} &= y_{4+0} = -2 \wedge y_{k_0+nd,2} = y_{6+0} = -1 \\ P(n_d = 1): y_{k_0+nd,1} &= y_{4+1} = 1 \wedge y_{k_0+nd,2} = y_{6+1} = -2 \end{aligned}$$

Basierend auf diesen ähnlichsten Werten lässt sich für jede Verschiebung ein Ausgang \hat{y}_{k_1+nd} schätzen. Hier gilt:

$$\begin{aligned} P(n_d = -1): \hat{y}_{k_1+nd} &= \frac{y_{k_0+nd,1} + y_{k_0+nd,2}}{2} = \frac{-1+1}{2} = 0 \\ P(n_d = 0): \hat{y}_{k_1+nd} &= \frac{y_{k_0+nd,1} + y_{k_0+nd,2}}{2} = \frac{-2-1}{2} = -1.5 \\ P(n_d = 1): \hat{y}_{k_1+nd} &= \frac{y_{k_0+nd,1} + y_{k_0+nd,2}}{2} = \frac{1+(-2)}{2} = -0.5 \end{aligned}$$

Um nun y_{k_0+nd} und \hat{y}_{k_1+nd} zu vergleichen, müssen die y_{k_0+nd} Werte für jede Verschiebung bestimmt werden. Hier gilt:

$$\begin{aligned} P(n_d = -1): y_{k_0+nd} &= y_{2-1} = 2 \\ P(n_d = 0): y_{k_0+nd} &= y_{2+0} = 4 \\ P(n_d = 1): y_{k_0+nd} &= y_{2+1} = -1 \end{aligned}$$

Die Differenz zwischen y_{k_0+nd} und \hat{y}_{k_1+nd} ergibt sich hier zu:

$$\begin{aligned} P(n_d = -1): \hat{y}_{k_1+nd} - y_{k_0+nd} &= 0 - 2 = -2 \\ P(n_d = 0): \hat{y}_{k_1+nd} - y_{k_0+nd} &= -1.5 - 4 = -5.5 \\ P(n_d = 1): \hat{y}_{k_1+nd} - y_{k_0+nd} &= -0.5 - (-1) = 0.5 \end{aligned}$$

Die grafische Darstellung der gezeigten Rechnungen ist in Abbildung 2.7 dargestellt. Berücksichtigt man die Tatsache, dass y_{k_0+nd} und \hat{y}_{k_1+nd} für $n_d = d$ identisch sein sollen, ist $\hat{d} = 1$ die beste Schätzung für die Totzeit zwischen u und y ist (da für $P(n_d = 1)$ die Differenz $\hat{y}_{k_1+nd} - y_{k_0+nd}$ minimal ist).

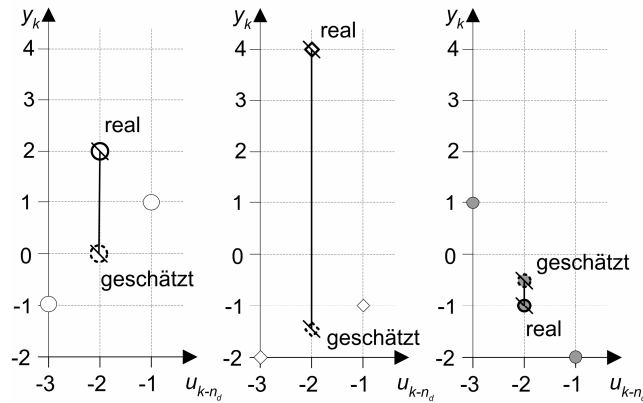


Abbildung 2.7: Differenz zwischen \hat{y}_{k_1+nd} und y_{k_0+nd} abhängig von der Verschiebung n_d zum Zeitpunkt $k_0=2$ ($n_d = -1$: \circ , $n_d = 0$: \diamond , $n_d = 1$: \bullet) [74]

Um eine verlässliche Schätzung der Totzeit basierend auf der Konsistenz zu ermöglichen, muss diese Prozedur für jeden Abtastpunkt, sofern $n \geq k - n_d > 0$ erfüllt ist, durchgeführt werden. Es muss daher ein automatisierter Algorithmus herangezogen werden. Es wird hier vorgeschlagen den Algorithmus des „Ersetzen fehlender Messwerte basierend auf den k nächsten Nachbarn“ zu nutzen. Diese Methode wurde entwickelt, um in einem mehrdimensionalen Datensatz fehlende Werte in einer Dimension durch die k ähnlichsten Werte (bezogen auf die vorhandenen Dimensionen), die so genannten nächsten Nachbarn, zu ersetzen. Dieser Vorgang ist für $k = 1$ in Abbildung 2.8 dargestellt. Die theoretischen Grundlagen dieses Algorithmus sind in [62] zu finden.

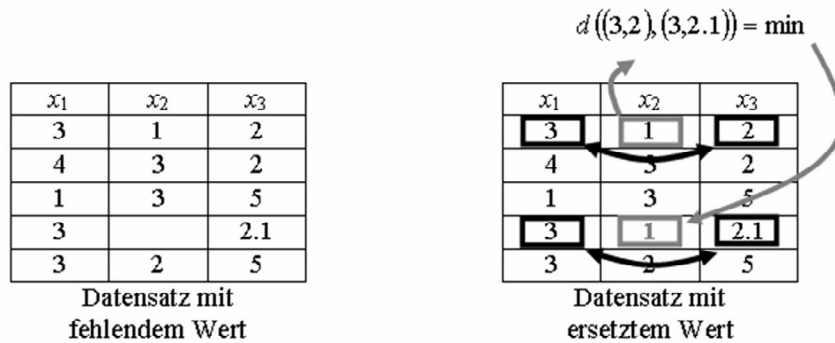


Abbildung 2.8: Ersetzung von fehlenden Messwerten mittels k nächster Nachbarn ($k = 1$)

Um Verwechslungen mit der diskreten Zeit k zu vermeiden, wird die Anzahl der nächsten Nachbarn im Folgenden mit k_{imp} bezeichnet. Die automatische Schätzung der Konsistenz für eine diskrete Verschiebung n_d unter Nutzung des gezeigten Algorithmus, basiert auf der Tatsache, aus den Punktmengen $P(n_d) = \{(u_{k-n_d}, y_k) \mid k = 1, 2, \dots, n\}$ nacheinander jeden Ausgang y_k wegzulassen und mit der genannten Routine erneut zu schätzen. Somit ist die quadratische Abweichungssumme

$$s(n_d) = \sum_{k=1}^n (\hat{y}_k - y_k)^2 \quad (2.18)$$

ein Maß für Konsistenz. Bedingt durch die Tatsache, dass diese nicht normiert ist, lassen sich Ergebnisse zweier Totzeitschätzungen nicht vergleichen, was wiederum eine Validierung unmöglich macht. Daher wird ein normierter Index vorgeschlagen gemäß

$$\rho(n_d) = \left(1 + \sqrt{\frac{s}{n} \cdot \frac{k_{imp}}{k_{imp} + 1}} \right)^{-1}. \quad (2.19)$$

Dieser erfüllt alle Anforderungen zur Automatisierbarkeit, da er zwischen 0 ($s = \infty$) und 1 ($s = 0$) liegt und somit normiert ist. Außerdem ist dessen Mittelwert unabhängig von der Parameterwahl k_{imp} und n . Die mathematische Begründung hierfür sei im Folgenden für zwei Zufallszahlen u_k und y_k allgemein gezeigt. Es gelte, dass u_k und y_k i.i.d. (*independent and identically distributed*, unabhängig und gleichartig verteilt) mit $u_k \sim N(0,1)$.

Das bedeutet für die nächsten Nachbarn $(y_{k,1}, y_{k,2}, \dots, y_{k,k_{imp}})$, dass $y_{k,1}, y_{k,2}, \dots, y_{k,k_{imp}} \sim N(0,1)$ und bedingt durch $E(y_{k,i}) = 0$ auch $E(y_{k,i}^2) = \sigma^2(y) = 1$

Der k -te geschätzte y -Wert (\hat{y}_k) ist demzufolge definiert durch

$$\hat{y}_k = \frac{1}{k_{imp}} \sum_{i=1}^{k_{imp}} y_{k,i} = \frac{y_{k,1}}{k_{imp}} + \dots + \frac{y_{k,k_{imp}}}{k_{imp}}. \quad (2.20)$$

wobei gilt, dass

$$(\hat{y}_k - y_k)^2 = \left(\underbrace{\frac{y_{k,1}}{k_{imp}} + \frac{y_{k,2}}{k_{imp}} + \dots + \frac{y_{k,k_{imp}}}{k_{imp}}}_{\hat{y}_k} - y_k \right)^2. \quad (2.21)$$

Unter der Nutzung der Vereinfachung $s := \sum_{i=1}^n ((\hat{y}_k - y_{ki})^2)$ lässt sich der zugehörige Erwartungswert herleiten gemäß

$$\begin{aligned} E((\hat{y}_k - y_k)^2) &= E(\hat{y}_k^2) - \underbrace{2E(\hat{y}_k \cdot y_k)}_{=0} + E(y_k^2) = \frac{1}{k_{imp}} \cdot \sigma^2(y) + \sigma^2(y) = \frac{k_{imp} + 1}{k_{imp}} \\ \Rightarrow E(s) &= E\left(\sum_{i=1}^n ((\hat{y}_k - y_{ki})^2)\right) = n \cdot E((\hat{y}_k - y_k)^2) = n \cdot \frac{k_{imp} + 1}{k_{imp}}. \end{aligned} \quad (2.22)$$

Um daraus die Verteilung von $\rho(n_d)$ ableiten zu können, wird die Tatsache genutzt, dass, sofern $X \sim N(\mu, \sigma^2)$ auch gilt $f_{lin}(X) \sim N(\mu_f, \sigma_f^2)$, wenn f_{lin} eine lineare Funktion der Zufallszahl X ist. Nichtlineare Funktionen lassen sich in der Umgebung eines Werts X_0 (im Folgenden zweckmäßig $x_0 = E(X)$) linearisieren, sofern die Umgebung nicht zu breit ist. Die schmale Umgebung kann etwa durch eine geringe

Varianz der Zufallszahl X erfüllt werden. Es gilt dann für die linearisierte Funktion unter Kenntnis von $E(X)$, dass

$$f_{lin}(X) = f(x_0) + \left. \frac{df(X)}{dX} \right|_{X=x_0} \cdot (X - x_0) = mX + b. \quad (2.23)$$

Varianz σ_f^2 und Erwartungswert μ_f lassen sich mithilfe der linearen Approximation bestimmen gemäß

$$\mu_f = E(f(X)) \approx E(f_{lin}(X)) = m \cdot E(X) + b \quad (2.24)$$

und

$$\begin{aligned} \sigma_f^2 &= E((f(X) - E(f(X)))^2) \approx E((f_{lin}(X) - E(f_{lin}(X)))^2) = \\ &= E((mX + b - (m \cdot E(X) + b))^2) \end{aligned} \quad (2.25)$$

Unter Kenntnis dieser Linearisierung lässt sich für $x_0 = E(s) = n \cdot \frac{k_{imp} + 1}{k_{imp}}$ mit $m = -\frac{1}{8} \cdot \left(\frac{1}{n} \cdot \frac{k_{imp}}{k_{imp} + 1} \right)$ und $b = \frac{5}{8}$ ableiten, sofern $\sigma^2(s) \ll 1$ angenommen wird, dass

$$E(\rho) = \frac{1}{2}. \quad (2.26)$$

Zwei unabhängige und normal verteilte Zeitreihen u_k und y_k haben somit einen erwarteten ρ -Wert von 0.5 und dieser ist somit unabhängig von n und k_{imp} .

Vergleichbar zu (2.12) lässt sich die Totzeit schätzen als

$$\hat{d} = \arg \max_{n_d \in d_{\min}, \dots, d_{\max}} \rho(n_d). \quad (2.27)$$

Der Algorithmus (KNNK, k-nächste-Nachbarn-Konsistenz) zur Berechnung von $\rho(n_d)$ unter Nutzung einer fertigen Routine zur *knn*-Ersetzung (bspw. *ce.impute* aus dem Paket *dprep* der mathematischen Software *R*) ist folgend aufgelistet.

Algorithmus KNNK: Bestimmung von $\rho(n_d)$

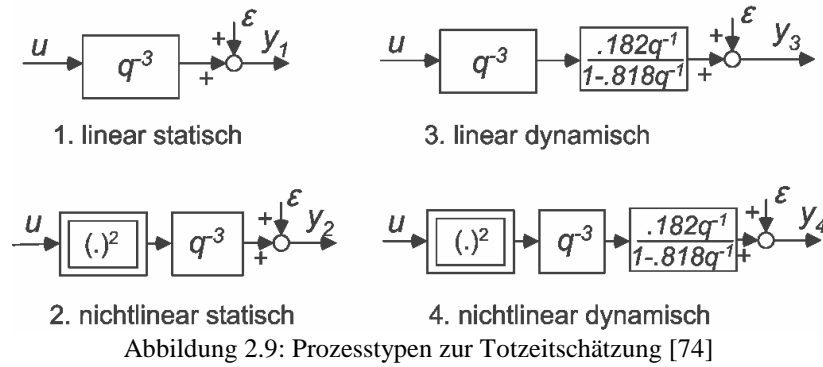
Der Startwert von n_d sei um eins kleiner als die minimal angenommene Totzeit d_{min} .

1. Inkrementiere n_d um 1.
2. Verschiebe den angenommenen Eingangsvektor u um n_d und erzeuge die Matrix $[u_{k-nd} \quad y_k]$.
3. Setze den Zählindex k und die Variable $s(n_d)$ zu Null.
4. Inkrementiere k um eins.
5. Speichere y_k separat und lösche es in der Matrix $\rightarrow [u_{k-nd} \quad NA]$
6. Ersetze den fehlenden y_k Wert durch die knn Routine und erzeuge somit \hat{y}_k .
7. Aktualisiere $s(n_d)$ gemäß: $s(n_d) = s(n_d) + (\hat{y}_k - y_k)^2$
8. Ersetze den geschätzten Wert \hat{y}_k durch den gespeicherten Wert y_k und gehe zu Schritt 4, bis k die Länge des Vektors erreicht hat.
9. Berechne $\rho(n_d) = \left(1 + \sqrt{\frac{s(n_d)}{n} \cdot \frac{k_{imp}}{k_{imp} + 1}}\right)^{-1}$
10. Gehe zu Schritt 1 und wiederhole die Berechnung bis n_d die maximal anzunehmende Totzeit d_{max} erreicht hat.

Es wurde gezeigt, dass die Totzeit zwischen zwei Signalen mittels Konsistenz bestimmt werden kann. Die Wirksamkeit dieser Methode wurde in [69] gezeigt und als „bislang vielversprechendstes Verfahren“ zur Totzeitschätzung dargestellt. Dieses Verfahren ist jedoch lediglich für statische Systeme anwendbar, bei denen der aktuelle Ausgang lediglich von einem vorangegangenen Eingang abhängt. Für dynamische Systeme p -ter Ordnung gilt dies nicht mehr. Hier hängt der aktuelle Ausgang von einem vorherigen Eingang und den letzten p Ausgängen ab. Dies führt dazu, dass die Schätzungen für \hat{y}_k drastisch schlechter werden und somit $\rho(n_d) \ll 1$ auch für den Fall, dass $n_d = d$. Bedingt durch die Tatsache, dass die Totzeit nach (2.27) allerdings lediglich nur auf dem Maximum der Funktion $\rho(n_d)$ basiert, wird in den meisten Fällen die Schätzung dennoch ein korrektes Ergebnis liefern. Lediglich Prozesse mit sehr kleinen Abtastzeiten oder Prozessen mit einer sehr großen Ordnung werden dazu führen, dass $\rho(n_d) \approx 0.5$. Dies macht eine korrekte Schätzung der Totzeit unmöglich, da $E(\rho(n_d)) = 0.5$. Für diesen Fall, muss ein modifizierter MISO-Ansatz genutzt werden, der am Ende dieses Kapitels gezeigt wird.

Um die Anwendbarkeit der Totzeitschätzung, basierend auf Konsistenzanalyse, auch für dynamische Prozesse zu zeigen, sei die Totzeit für die folgenden vier Prozesse zu schätzen. Die Prozesse sind dabei:

- linear statisch,
- nichtlinear statisch,
- linear dynamisch,
- nichtlinear dynamisch (Hammerstein), siehe Abbildung 2.9.



Alle Prozesse haben als Eingang $u(k) \sim N(0,9)$ und das zusätzliche weiße Rauschen $\varepsilon(k) \sim N(0,0.01)$. Der Verlauf der vier simulierten Ausgänge ($y_{1,k}, \dots, y_{4,k}$) mit den zugehörigen Totzeitschätzungen ist in Abbildung 2.10 dargestellt. Der angenommene Eingang ist in allen Simulationen stets u .

Es fällt auf, dass die Totzeit für jeden Prozess, auch in der Gegenwart von weißem Rauschen, korrekt geschätzt wird. Wie bereits beschrieben, sind die maximalen $\rho(n_d)$ Werte für die dynamischen Simulationen kleiner (~ 0.56) als für die statischen Simulationen (~ 0.9).

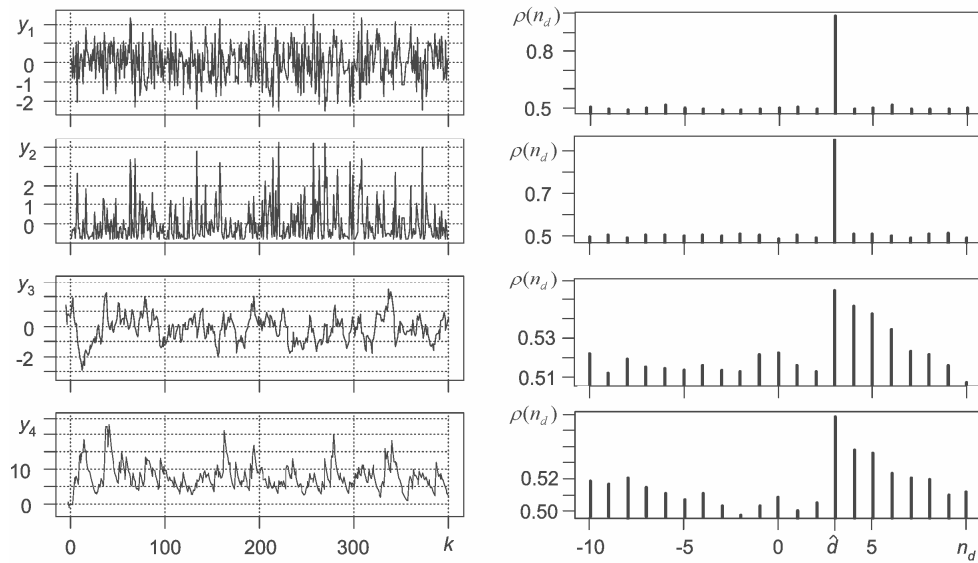


Abbildung 2.10: Zeitverlauf der vier Prozessstypen (links) mit den zugehörigen Totzeitschätzungen (rechts) [74]

Totzeitschätzung für MISO Systeme basierend auf KNNK

Es wurde bisher gezeigt, dass das neue Verfahren zur Totzeitschätzung ebenfalls und ohne Modifikation für nichtlineare Systeme anwendbar ist. Neben dieser Tatsache ist ein weiterer Vorteil, dass es ebenfalls für MISO-Systeme mit M_u Eingängen anwendbar ist, bei denen jeder Eingang eine unterschiedliche Totzeit bezogen auf den Ausgang besitzen kann. In diesem Fall muss lediglich die Punktmenge erweitert werden, gemäß

$$P(n_{d,u1}, \dots, n_{d,uM_u}) = \left\{ \left(u_{1,k-n_{d,u1}}, u_{2,k-n_{d,u2}}, \dots, u_{M_u,k-n_{d,uM_u}}, y_k \right) \right\}, \quad (2.28)$$

wobei $n_{d,ui}$ die diskrete Verschiebung des i -ten angenommenen Eingangs angibt. Es gelten dieselben Einschränkungen wie für den einfachen SISO-Fall.

Für diesen Fall muss der KNNK-Algorithmus leicht modifiziert werden. Anstatt lediglich einen angenommenen Eingang u um n_d zu verschieben, müssen für diesen Fall alle M_u angenommenen Eingänge (u_1, u_2, \dots, u_{M_u}) separat um $n_{d,u1}, \dots, n_{d,uM_u}$ verschoben werden. Dadurch wird $\rho(\cdot)$ eine Funktion von $n_{d,u1}, \dots, n_{d,uM_u}$, was wiederum zu einer (M_u+1) dimensional Visualisierung führt.

Die Ergebnisse einer MISO-Totzeitschätzung seien für einen einfachen MISO-Prozess gezeigt, mit

$$y_k = u_{1,k-3} + u_{2,k+4}, \quad (2.29)$$

Das Ergebnis der Totzeitschätzung mit den angenommenen Eingängen u_1 und u_2 ist in Abbildung 2.11 dargestellt.

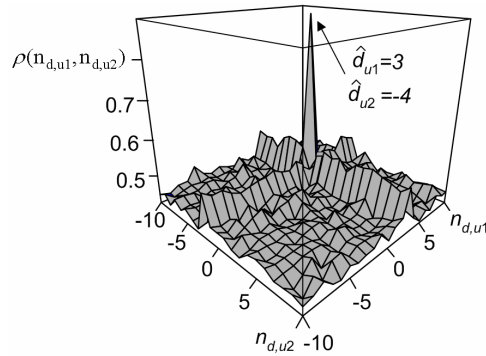


Abbildung 2.11: Totzeitschätzung für MISO-Prozess mit zwei unterschiedlichen Totzeiten $d_{u1} = 3$ und $d_{u2} = -4$ [74]

Anmerkung: Der Prozess, welcher in (2.29) dargestellt ist, besitzt eine negative Totzeit ($d = -4$). Dies scheint widersprüchlich, wird die Gleichung jedoch unter Nutzung des Verschiebeoperators q^{-i} wie folgt umgeschrieben, so wird klar, dass der eigentliche Ausgang $u_{2,k}$ ist:

$$\begin{aligned} y_k &= u_{1,k-3} + u_{2,k+4} \\ \Leftrightarrow y_k &= u_1 q^{-3} + u_2 q^4 \Big| \cdot q^{-4} \\ \Leftrightarrow y_k q^{-4} &= u_1 q^{-7} + u_2 \\ \Leftrightarrow u_2 &= y_k q^{-4} - u_1 q^{-7} \\ \Leftrightarrow u_{2,k} &= y_{k-4} - u_{1,k-7} \end{aligned}$$

Ähnliches wurde bereits in 2.2.1 für die Kreuzkorrelation gezeigt, bei der eine negative Totzeitschätzung dazu führt, dass die Annahme des Eingangs falsch ist. Gleiches gilt, wie gezeigt, auch hier.

Validierung der Totzeitschätzung basierend auf KNNK

Ziel der Validierung ist es, festgelegte Grenzwerte zu finden, die Aussagen über die Qualität und Wiederholbarkeit der durchgeführten Totzeitschätzung ermöglichen. Oder

mit anderen Worten: Was ist der minimale $\rho(\hat{d})$ Wert, unterhalb dessen die Schätzung für \hat{d} verworfen werden sollte?

Diese Fragestellung kann auf ein mathematisches Problem reduziert werden, sofern die Verteilung von $\rho(n_d)$ im Falle von zwei unabhängigen, skalierten normalverteilten Zufallszahlen u_k und y_k bekannt ist.

Unter Nutzung der bereits gezeigten Linearisierung lässt sich die Verteilung von $\rho(n_d)$ bestimmen. Es gilt zunächst, dass

$$\begin{aligned} E(\hat{y}_k) &= E\left(\frac{1}{k_{imp}} \sum_{i=1}^{k_{imp}} y_{k,i}\right) = 0 \\ \Rightarrow E((\hat{y}_k - E(\hat{y}_k))^2) &= E(\hat{y}_k^2) = \frac{1}{k_{imp}} \end{aligned} \quad (2.30)$$

Für die Verteilung gilt, dass

$$\hat{y}_k \sim N(0, 1/k_{imp}). \quad (2.31)$$

Und somit gilt für die Verteilung der Abweichung $\hat{y}_k - y_k \sim N(0, \sigma^2)$ mit der unbekannten Varianz σ^2 . Unter Beachtung der Definition der Chi-Quadrat (χ^2) Verteilung (z.B. [16]) kann gesagt werden, dass $(\hat{y}_k - y_k)^2 \sim \chi^2(\mu_{\chi^2})$, wobei μ_{χ^2} die Anzahl an Freiheitsgraden ist. Für eine Chi-Quadrat verteilte Zufallsvariable X mit μ_{χ^2} gilt, dass für den Erwartungswert von X gilt

$$E(X) = \mu_{\chi^2}. \quad (2.32)$$

Unter Nutzung der gemachten Erkenntnisse hinsichtlich des Erwartungswerts von $(\hat{y}_k - y_k)^2$ gilt somit für μ_{χ^2} , dass

$$\mu_{\chi^2} = E((\hat{y}_k - y_k)^2) = \frac{k_{imp} + 1}{k_{imp}} \Rightarrow (\hat{y}_k - y_k)^2 \sim \chi^2\left(\frac{k_{imp} + 1}{k_{imp}}\right). \quad (2.33)$$

Wird erneut die Abkürzung $s := \sum_{k=1}^n (\hat{y}_k - y_k)^2$ verwendet, so gilt für die Verteilung von s , dass

$$s \sim \chi^2\left(n \cdot \frac{k_{imp} + 1}{k_{imp}}\right). \quad (2.34)$$

Durch Einführung des Parameters q folgt ferner, dass

$$q := \frac{s}{n} \cdot \frac{k_{imp}}{k_{imp} + 1} \Rightarrow \rho = \frac{1}{1 + \sqrt{q}}. \quad (2.35)$$

Für eine große Anzahl an Messungen n ($n > 100$) ist gezeigt (z.B. durch [16]), dass

$$q \sim N(1, 2f), \quad f := \frac{k_{imp}}{n \cdot (k_{imp} + 1)}. \quad (2.36)$$

Durch die Einführung des Parameters p lässt sich der Term weiter vereinfachen gemäß

$$p := \sqrt{q} \Rightarrow \rho = \frac{1}{1+p}. \quad (2.37)$$

Die Verteilung von p lässt sich ebenfalls mithilfe der gezeigten Linearisierung ableiten. Es gilt, dass $E(q) = 1$. Die Gültigkeit der Linearisierung wird gewährleistet, da angenommen werden kann, dass $\sigma_q^2 \ll 1$. Demzufolge gilt, dass

$$p \sim N\left(1, \frac{1}{2}f\right). \quad (2.38)$$

Mit $r := 1 + p \Rightarrow \rho = \frac{1}{r}$ gilt ohne Beweis, dass

$$r \sim N\left(2, \frac{1}{2}f\right) \quad (2.39)$$

und in Zusammenfassung $\rho = \frac{1}{r}$ gilt, erneut unter Nutzung der Linearisierung, dass

$$\rho \sim N\left(\frac{1}{2}, \frac{1}{32} \cdot f\right) \text{ oder } \rho \sim N\left(\frac{1}{2}, \frac{1}{32} \cdot \frac{k_{imp}}{n \cdot (k_{imp} + 1)}\right). \quad (2.40)$$

Der Einfluss von n und k_{imp} auf die Verteilung von $\rho(n_d)$ ist in Abbildung 2.12 dargestellt. Es ist offensichtlich, dass der Einfluss von k_{imp} vernachlässigbar ist, da der Quotient $k_{imp}/(k_{imp} + 1)$ für große k_{imp} asymptotisch gegen eins geht.

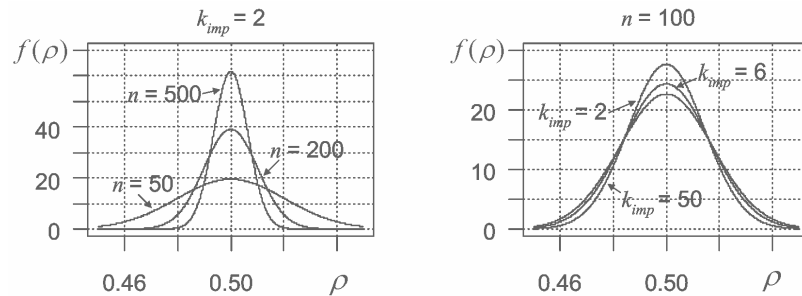


Abbildung 2.12: Einfluss der Anzahl an Messungen n und die Anzahl der nächsten Nachbarn k_{imp} auf die Verteilung von ρ [74]

Unter Kenntnis der Verteilung und unter Annahme eines Signifikanzlevels α von 0.0001 lässt sich beispielhaft folgende einseitige Nullhypothese aufstellen:

Angenommen n sei 200 und k_{imp} sei 2, so sind mit einer Wahrscheinlichkeit von 0.9999 alle $\rho(\cdot)$ Werte kleiner als 0.534 für zwei unabhängige, skalierte, normalverteilte Zeitreihen.

Dies führt im Gegenschluss zu folgender Alternativhypothese:

Angenommen n sei 200, k_{imp} sei 2 und sei $\rho(n_d)$ größer als $\rho_{crit} = 0.534$, so kann mit einer Wahrscheinlichkeit von 0.9999 angenommen werden, dass die Messreihen, unter der Voraussetzung, dass beide Messreihen skaliert und normalverteilt sind, nicht unabhängig sind.

Die Voraussetzung der Normalverteilung ist oft gegeben oder wird zumindest in guter Näherung erreicht. Dies führt bspw. für $n = 200$ und $k_{imp} = 2$ dazu, dass ρ Werte für einen Zeitverzug n_d , welche größer sind als $\rho_{crit} = 0.534$ als verlässliche Schätzung der Totzeit angenommen werden können, da ein gleiches Ergebnis nicht zufällig hätte erreicht werden können.

Diese Kenntnis von ρ_{crit} kann somit in die Schätzung der Totzeit \hat{d} integriert werden mit

$$\hat{d} = \begin{cases} \arg \max_{n_d \in d_{\min}, \dots, d_{\max}} \rho(n_d), & \text{if } \rho(n_d) > \rho_{crit} \\ \emptyset, & \text{else} \end{cases} \quad (2.41)$$

Anmerkung: Die Hauptidee der Totzeitschätzung basierend auf knn ist weiterhin für alle generellen Messdaten möglich, auch wenn diese nicht einmal in guter Näherung normalverteilt sind (*typical operating data*, typische Betriebsdaten). Für diesen Fall ist lediglich der Vergleich mit ρ_{crit} nicht möglich.

Um die Integration von (2.41) für die Totzeitschätzung zu zeigen, ist im Folgenden noch einmal das Ergebnis der Totzeitschätzungen aus Abbildung 2.9 und Abbildung 2.11 gezeigt. Der kritische Wert ρ_{crit} ist durch eine gestrichelte Linie dargestellt. Alle ρ Werte kleiner als ρ_{crit} sind grau dargestellt, die restlichen schwarz.

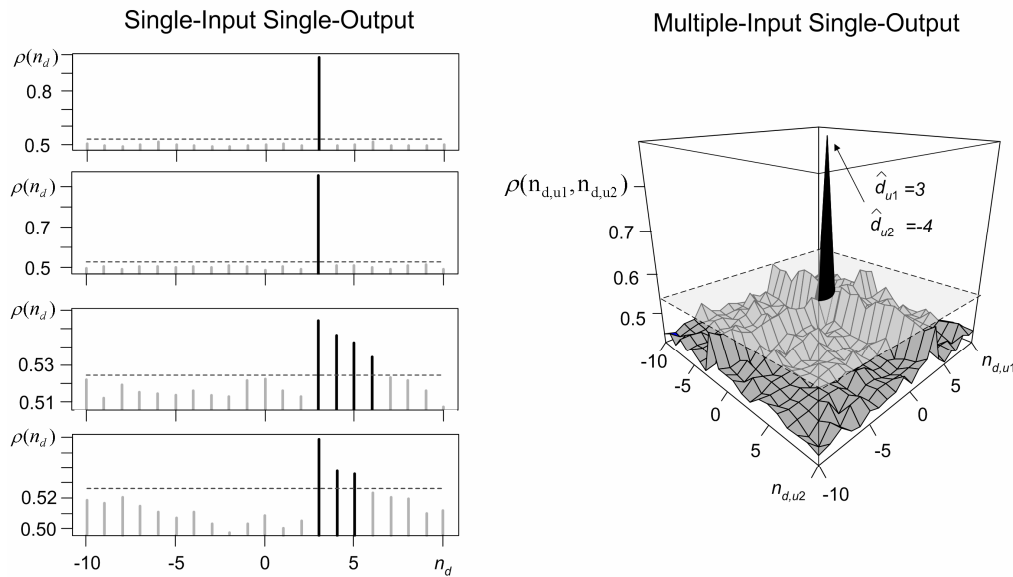


Abbildung 2.13: Totzeitschätzung für die Prozesse aus Abbildung 4.10 und Abbildung 2.11 mit der dargestellten Signifikanzgrenze ρ_{crit} [74]

Totzeitschätzung für Prozesse hoher Ordnung basierend auf KNNK

Betrachtet man die Ergebnisse aus Abbildung 2.13, so ist ersichtlich, dass die Schätzung aller Totzeiten weiterhin möglich ist, allerdings liegen für den dynamischen Fall die $\rho(d)$ Werte nur noch knapp oberhalb der Signifikanzgrenze. Höhere Prozessordnungen oder kleinere Abtastzeiten führen dazu, dass die $\rho(d)$ Werte unterhalb von ρ_{crit} liegen. Eine Möglichkeit, dies zu vermeiden ist die Anwendung eines abgewandelten MISO-Ansatzes für einen Prozess p -ter Modellordnung, gemäß

$$P(n_d) = \left\{ (u_{k-n_d}, y_{k-p}, y_{k-(p-1)}, \dots, y_{k-1}, y_k) \right\}. \quad (2.42)$$

Die Visualisierung dieser $p+2$ dimensional Punktmenge bleibt jedoch zwei-dimensional, da sie lediglich von n_d abhängt und ist somit vergleichbar mit dem einfachen SISO-Ansatz. Um dies zu zeigen, ist die Totzeit für y_3 und y_4 aus Abbildung 2.13 unter der Annahme erneut durchgeführt, dass es sich um einen Prozess erster Ordnung handelt, was zu der Punktmenge $P(n_d) = \left\{ (u_{k-n_d}, y_{k-1}, y_k) \right\}$ führt. Die Ergebnisse sind in Abbildung 2.14 dargestellt. Man kann eindeutig die Verbesserung erkennen.

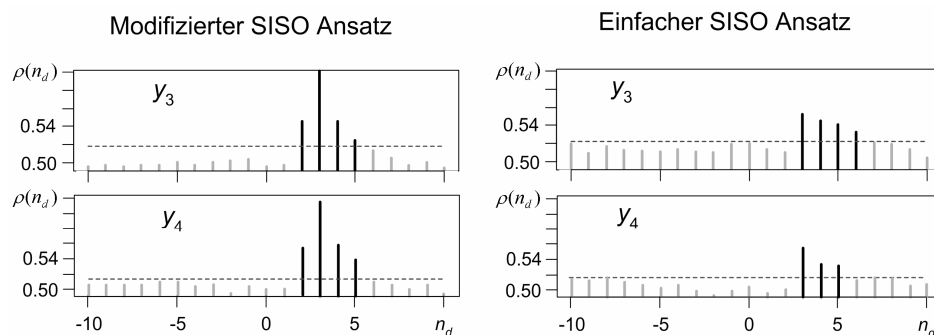


Abbildung 2.14: Totzeitschätzung für dynamische Prozesse mit abgewandelten MISO-Ansatz (links) und dem einfachen SISO-Ansatz (rechts) [74]

Anmerkung: Auf den ersten Blick scheint die a priori Schätzung der Modellordnung ein Widerspruch zu 1.2 zu sein, allerdings gilt ohne Beweis, jedoch basierend auf der *knn*-Theorie, dass es keine signifikante Rolle spielt, ob p größer als die eigentliche Prozessordnung gewählt wird. Um dies zu zeigen, wird erneut die Totzeit für y_3 und y_4 geschätzt, dieses mal jedoch mit dem Ansatz $p = 4$ und somit $P(n_d) = \{u_{k-n_d}, y_{k-4}, y_{k-3}, y_{k-2}, y_{k-1}, y_k\}$. Die Ergebnisse dieser Schätzung im Vergleich zur vorherigen Schätzung ($p = 1$) sind in Abbildung 2.15 dargestellt.

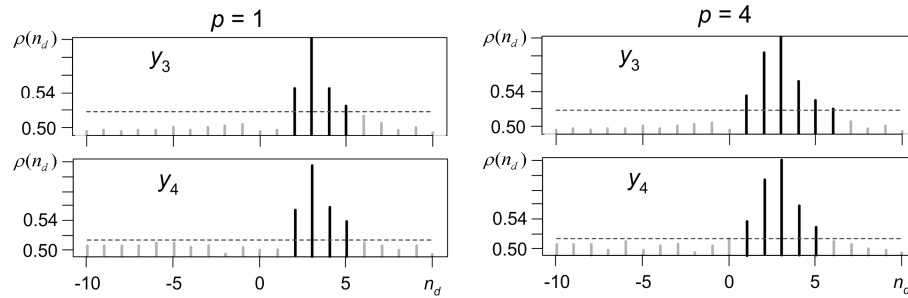


Abbildung 2.15: Totzeitschätzung für dynamische Prozesse mit $p = 1$ (links) und $p = 4$ (rechts) [74]

2.2.3 Simulationsergebnisse

Die Vorgehensweise zur Offline-Erkennung von Verursachern anlagenweiter Fehler basierend auf der Beeinflussungsmatrix wird in diesem Abschnitt an unterschiedlichen Simulationen gezeigt und getestet. Dabei sowohl unter Nutzung der vereinfachten Vorgehensweise im Falle von ausschließlich linearen Prozessen nach Abschnitt 2.2.1 und unter Nutzung des erweiterten Ansatzes auch im Falle von nichtlinearen Prozessen nach Abschnitt 2.2.2. Zum Testen sei dafür der Benchmarkprozess aus Abbildung 2.16 gegeben.

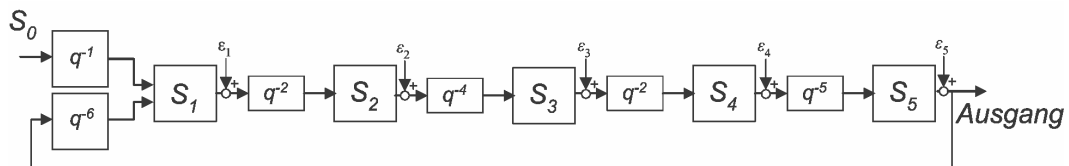


Abbildung 2.16: Benchmarkprozess zur Identifikation von anlagenweiten Fehlerverursachern [74]

Sofern nichts anderes angegeben wird, besteht der Benchmarkprozess aus fünf linearen, statischen Teilprozessen S_i der Form $y_i = a_0 + a_1 \cdot u_i$. Der Ausgang aller Teilprozesse wird durch weißes Rauschen mit dem Mittelwert 0 und der Standardabweichung 0.1 gestört. Der Prozesseingang S_0 sei nicht messbar und besteht aus einer normalverteilten Zufallszahl mit dem Mittelwert 1 und der Standardabweichung 1.

Simulationsbeispiel 2.2: Lineare Teilprozesse, normaler Betrieb, beeinflusst durch S_0

In dieser Simulation arbeitet die Anlage normal und wird hauptsächlich durch die externen Störungen von S_0 beeinflusst. Die Zeitverläufe der einzelnen Messungen sind in Abbildung 2.17 dargestellt.

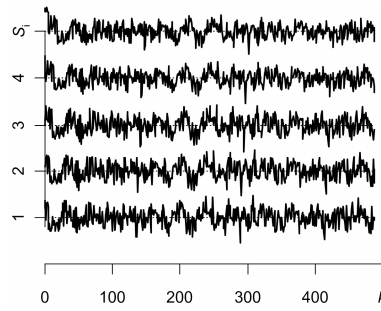


Abbildung 2.17: Skalierte Zeitverläufe der Teilprozesse S_1 bis S_5 im Falle der normalen Betriebsweise [74]

Basierend auf der Vorgehensweise, welche in Abbildung 2.2 dargestellt ist, muss zunächst die Beeinflussungsmatrix aufgestellt werden. Dies kann, wie gezeigt wurde, auf zwei Arten gemacht werden. Mittels Kreuzkorrelation (\mathbf{Z}_{ccf}) und mittels KNNK-Algorithmus (\mathbf{Z}_{knn}). Beide Matrizen sind in diesem Fall identisch mit

$$\mathbf{Z}_{ccf} = \mathbf{Z}_{knn} = \begin{matrix} & \begin{matrix} S_1 & S_2 & S_3 & S_4 & S_5 \end{matrix} \\ \begin{matrix} S_1 \\ S_2 \\ S_3 \\ S_4 \\ S_5 \end{matrix} & \begin{bmatrix} - & 1 & 1 & 1 & 1 \\ 1 & - & 2 & 2 & 2 \\ 1 & 2 & - & 3 & 3 \\ 1 & 2 & 3 & - & 4 \\ 1 & 2 & 3 & 4 & - \end{bmatrix} \end{matrix}. \quad (2.43)$$

Bedingt durch die Tatsache, dass S_1 der identifizierte Verursacher ist und unter der Voraussetzung, dass S_0 nicht messbar ist, muss eine ausführliche Fehlerdiagnose in S_1 durchgeführt werden. Für den Fall, dass die Anlage normal arbeitet, wird im Prozess kein Fehler auffindbar sein.

Simulationsbeispiel 2.3: Lineare Teilprozesse mit einer Oszillation in S_3

Im Zusatz zu der Anlage aus Abbildung 2.16 wird in dieser Simulation eine stabile Oszillation mit der Periodendauer 20π zum Ausgang des Teilprozesses S_3 addiert. Die Zeitverläufe aller Teilprozesse sind in Abbildung 2.18 gezeigt.

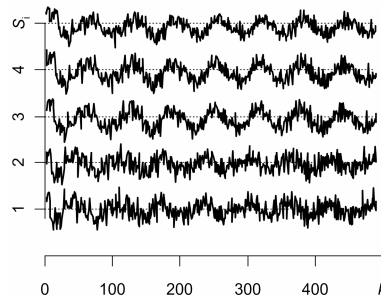


Abbildung 2.18: Skalierte Zeitverläufe der Teilprozesse S_1 bis S_5 im Falle einer Oszillation in S_3 [74]

Die daraus resultierenden Beeinflussungsmatrizen \mathbf{Z}_{ccf} und \mathbf{Z}_{knn} sind folgend gegeben, durch

$$\mathbf{Z}_{ccf} = \mathbf{Z}_{knn} = \begin{matrix} & S_1 & S_2 & S_3 & S_4 & S_5 \\ \begin{matrix} S_1 \\ S_2 \\ S_3 \\ S_4 \\ S_5 \end{matrix} & \begin{bmatrix} - & 1 & 3 & 4 & 5 \\ 1 & - & 3 & 4 & 5 \\ 3 & 3 & - & 3 & 3 \\ 4 & 4 & 3 & - & 4 \\ 5 & 5 & 3 & 4 & - \end{bmatrix} \end{matrix}. \quad (2.44)$$

Da es hierbei einen identifizierten Verursacher gibt, welcher nicht S_1 ist, muss eine ausführliche Fehlerdiagnose in S_3 durchgeführt werden.

Simulationsbeispiel 2.4: Lineare Teilprozesse mit einer periodischen Störung in S_3

Im Zusatz zu der Anlage aus Abbildung 2.16 wird in dieser Simulation eine pulsartige periodische Störung (± 10) zum Ausgang des Teilprozesses S_3 addiert. Die Zeitverläufe aller Teilprozesse sind in Abbildung 2.19 gezeigt.

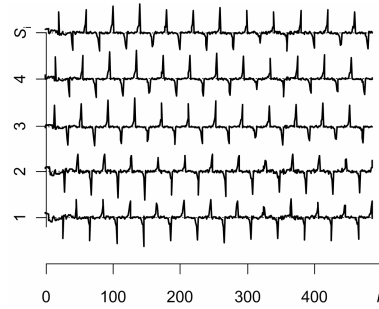


Abbildung 2.19: Skalierte Zeitverläufe der Teilprozesse S_1 bis S_5 im Falle einer Oszillation in S_3 [74]

Die daraus resultierenden Beeinflussungsmatrizen \mathbf{Z}_{ccf} und \mathbf{Z}_{knn} errechnen sich gemäß

$$\mathbf{Z}_{ccf} = \mathbf{Z}_{knn} = \begin{matrix} & S_1 & S_2 & S_3 & S_4 & S_5 \\ \begin{matrix} S_1 \\ S_2 \\ S_3 \\ S_4 \\ S_5 \end{matrix} & \begin{bmatrix} - & 1 & 3 & 4 & 5 \\ 1 & - & 3 & 4 & 5 \\ 3 & 3 & - & 3 & 3 \\ 4 & 4 & 3 & - & 4 \\ 5 & 5 & 3 & 4 & - \end{bmatrix} \end{matrix}. \quad (2.45)$$

Da es auch hierbei einen identifizierten Verursacher gibt, welcher nicht S_1 ist, muss eine ausführliche Fehlerdiagnose in S_3 durchgeführt werden.

Simulationsbeispiel 2.5: Lineare Teilprozesse mit einer arbiträren Störung in S_2

Im Zusatz zu der Anlage aus Abbildung 2.16 wird in dieser Simulation eine arbiträre Störung zum Ausgang des Teilprozesses S_2 addiert. Die Wahrscheinlichkeit für das Auftreten der Störung beträgt 5% und die Größe der Störung entspricht einer gleichverteilten Zufallszahl zwischen 1 und 5. Die Zeitverläufe aller Teilprozesse sind in Abbildung 2.20 gezeigt.

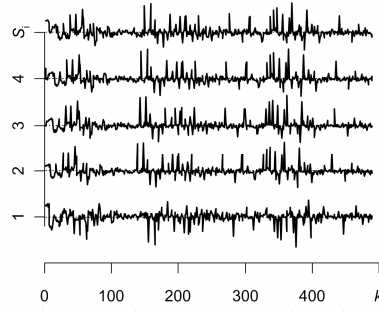


Abbildung 2.20: Skalierte Zeitverläufe der Teilprozesse S_1 bis S_5 im Falle einer arbiträren Störung in S_2 [74]

Die daraus resultierenden Beeinflussungsmatrizen \mathbf{Z}_{cf} und \mathbf{Z}_{kn} sind folgend gegeben durch

$$\mathbf{Z}_{cf} = \mathbf{Z}_{kn} = \begin{matrix} & \begin{matrix} S_1 & S_2 & S_3 & S_4 & S_5 \end{matrix} \\ \begin{matrix} S_1 \\ S_2 \\ S_3 \\ S_4 \\ S_5 \end{matrix} & \begin{bmatrix} - & 2 & 3 & 4 & 5 \\ 2 & - & 2 & 2 & 2 \\ 3 & 2 & - & 3 & 3 \\ 4 & 2 & 3 & - & 4 \\ 5 & 2 & 3 & 4 & - \end{bmatrix} \end{matrix}. \quad (2.46)$$

Der hier identifizierte Verursacher ist S_2 , was dazu führt, dass eine ausführliche Fehlerdiagnose in S_2 durchgeführt werden muss.

Simulationsbeispiel 2.6: Lineare Teilprozesse mit einer Oszillation in S_1

Diese Simulation ist vergleichbar mit Simulationsbeispiel 2.3, dieses Mal jedoch mit der Quelle in S_1 . Die Zeitverläufe aller Teilprozesse sind in Abbildung 2.21 gezeigt.

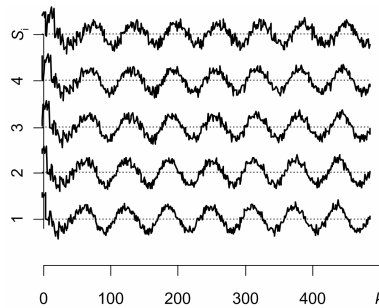


Abbildung 2.21: Skalierte Zeitverläufe der Teilprozesse S_1 bis S_5 im Falle einer Oszillation in S_1 [74]

Die daraus resultierenden Beeinflussungsmatrizen \mathbf{Z}_{cf} und \mathbf{Z}_{kn} sind folgend gegeben durch

$$\mathbf{Z}_{ccf} = \mathbf{Z}_{knn} = \begin{matrix} & S_1 & S_2 & S_3 & S_4 & S_5 \\ \begin{matrix} S_1 \\ S_2 \\ S_3 \\ S_4 \\ S_5 \end{matrix} & \begin{bmatrix} - & 1 & 1 & 1 & 1 \\ 1 & - & 2 & 2 & 2 \\ 1 & 2 & - & 3 & 3 \\ 1 & 2 & 3 & - & 4 \\ 1 & 2 & 3 & 4 & - \end{bmatrix} \end{matrix}. \quad (2.47)$$

Da es sich bei dem identifizierten Verursacher um S_1 handelt, muss dort eine ausführliche Diagnose durchgeführt werden.

Simulationsbeispiel 2.7: Lineare Teilprozesse mit einer Oszillation in S_3 und großen externen Störungen durch S_0

Um die Grenzen der Verfahrens nach Abschnitt 2.2 zu zeigen, ist in dieser Simulation erneut die Quelle einer Oszillation in S_3 zu finden, allerdings ist die Größe der externen Störungen in dieser Simulation größer ($S_0 \sim N(1,4)$). Die Zeitverläufe aller Teilprozesse sind in Abbildung 2.22 gezeigt.

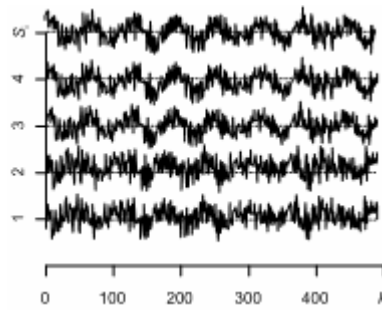


Abbildung 2.22: Skalierte Zeitverläufe der Teilprozesse S_1 bis S_5 im Falle einer Oszillation in S_3 und einer großen externen Störung [74]

Die daraus resultierenden Beeinflussungsmatrizen \mathbf{Z}_{ccf} und \mathbf{Z}_{knn} sind durch

$$\mathbf{Z}_{ccf} = \mathbf{Z}_{knn} = \begin{matrix} & S_1 & S_2 & S_3 & S_4 & S_5 \\ \begin{matrix} S_1 \\ S_2 \\ S_3 \\ S_4 \\ S_5 \end{matrix} & \begin{bmatrix} - & 1 & 1 & 1 & 1 \\ 1 & - & 2 & 2 & 2 \\ 1 & 2 & - & 3 & 3 \\ 1 & 2 & 3 & - & 4 \\ 1 & 2 & 3 & 4 & - \end{bmatrix} \end{matrix}. \quad (2.48)$$

gegeben.

Bedingt durch die Tatsache, dass zwar alle Teilprozesse durch S_3 beeinflusst werden, die maßgebliche Einflussgröße jedoch S_0 ist, wird in dieser Simulation fälschlicherweise S_1 als Verursacher identifiziert. Es lässt sich daraus folgende wichtige Einschränkung ableiten: Der Einfluss des Fehlers, hervorgerufen durch einen Teilprozess, muss größer sein, als der Einfluss der äußeren Störung. Dies scheint zwar eine Einschränkung zu sein, allerdings spielt hier die Dauer eine große Rolle. In der Praxis sind externe Störungen zwar häufig größer als die Störungen durch interne Fehler, jedoch dauern diese externen Störungen (z.B. Feedschwankungen) i.d.R. nicht so lange an wie interne

Störungen (z.B. falsch eingestellte Regler). Womit diese Tatsache keine signifikante Rolle für die praktische Anwendbarkeit spielt.

Simulationsbeispiel 2.8: Nichtlineare Teilprozesse mit einer Oszillation in S_4

Um die Unterschiede zwischen der Verfahren aus Abschnitt 2.2.1 und aus Abschnitt 2.2.2. zu zeigen, besteht die Anlage in dieser Simulation aus zwei nichtlinearen (quadratischen) statischen Teilprozessen (S_2 und S_3) und einer stabilen Oszillation in S_4 . Die Zeitverläufe aller Teilprozesse sind in Abbildung 2.23 gezeigt.

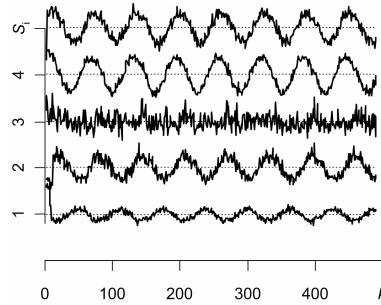


Abbildung 2.23: Skalierte Zeitverläufe der nichtlinearen Teilprozesse S_1 bis S_5 im Falle einer Oszillation in S_4 [74]

Die daraus resultierenden Beeinflussungsmatrizen \mathbf{Z}_{ccf} und \mathbf{Z}_{knn} unterscheiden sich nun und sind gegeben durch

$$\mathbf{Z}_{ccf} = \begin{matrix} & \begin{matrix} S_1 & S_2 & S_3 & S_4 & S_5 \end{matrix} \\ \begin{matrix} S_1 \\ S_2 \\ S_3 \\ S_4 \\ S_5 \end{matrix} & \begin{bmatrix} - & 1 & 3 & 4 & 5 \\ 1 & - & 2 & 4 & 5 \\ 3 & 2 & - & 3 & 5 \\ 4 & 4 & 3 & - & 4 \\ 5 & 5 & 5 & 4 & - \end{bmatrix} \end{matrix} \text{ und} \quad (2.49)$$

$$\mathbf{Z}_{knn} = \begin{matrix} & \begin{matrix} S_1 & S_2 & S_3 & S_4 & S_5 \end{matrix} \\ \begin{matrix} S_1 \\ S_2 \\ S_3 \\ S_4 \\ S_5 \end{matrix} & \begin{bmatrix} - & 1 & 1 & 4 & 5 \\ 1 & - & 2 & 4 & 5 \\ 1 & 2 & - & 4 & 5 \\ 4 & 4 & 4 & - & 4 \\ 5 & 5 & 5 & 4 & - \end{bmatrix} \end{matrix}.$$

Wie zu erwarten war, liefert lediglich die knn -Methode den korrekten Verursacher.

2.2.4 Laufzeitanalyse

Die in Abschnitt 2.2 dargestellten Verfahren eignen sich aufgrund ihres hohen Rechenaufwandes hauptsächlich für den Offline-Einsatz, vor allem bei zu geringen Abtastzeiten. Dies ist dadurch bedingt, dass bei M Teilprozessen $(M^2-M)/2$ Berechnungen zur Bestimmung der symmetrischen Beeinflussungsmatrix durchgeführt werden müssen. Die Anzahl an Berechnungen steigt somit quadratisch mit

zunehmender Anzahl an Teilprozessen. Es gilt somit generell für die Offline-Ansätze die Laufzeit

$$O(M^2). \quad (2.50)$$

Beide Offline-Verfahren unterscheiden sich noch einmal deutlich in ihrem Rechenaufwand. Das Verfahren aus Abschnitt 2.2.1, welches die Kreuzkorrelation nutzt, besitzt (bezogen auf die Anzahl an Messungen n) eine Laufzeit von

$$O(n). \quad (2.51)$$

Das Verfahren aus Abschnitt 2.2.2, welches auf dem KNNK-Algorithmus basiert, besitzt (bei $k_{imp} = \text{const}$) die Laufzeit

$$O(n^2). \quad (2.52)$$

Dies ist dadurch zu begründen, dass die Routine der *knn*-Ersetzung, welche die Laufzeit $O(n)$ besitzt, n -mal ausgeführt wird. Da die genannten Klassenzugehörigkeiten jedoch keinen Eindruck über die tatsächliche Rechendauer vermitteln, ist in Abbildung 2.24 die Rechendauer für die Totzeitschätzung zweier Vektoren der Länge n mittels Kreuzkorrelation und KNNK-Algorithmus ($k_{imp} = 2$) gezeigt. Die Spannweite beträgt in beiden Fällen $d_{min} = -1$, $d_{max} = 1$ und wurden auf einem 1.8 GHz AMD Rechner mit 1 GB RAM durchgeführt.

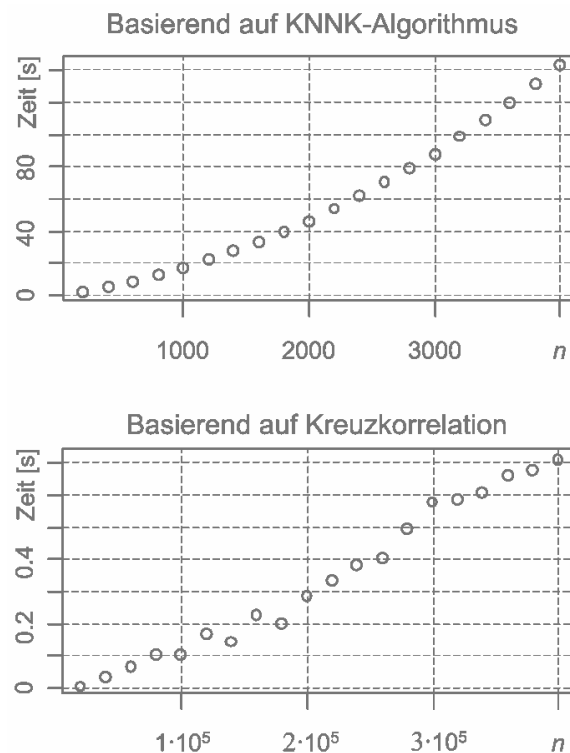


Abbildung 2.24: Rechenzeit in Abhängigkeit der Vektorlänge

2.3 Vereinfachte Erkennung im Online-Betrieb

In Abschnitt 2.2 wurde ein Offline-Verfahren zur Erkennung eines anlagenweiten Fehlerverursachers gezeigt. Der Offline-Einsatz bedingt sich durch den hohen Rechenaufwand und die damit verbundene Rechenzeit. In diesem Abschnitt wird ein Verfahren gezeigt, welches erheblich weniger Rechenzeit bedarf. Es basiert auf einem Standardverfahren der Signalanalyse, welches jedoch, wie im Folgenden gezeigt wird, auch für die Erkennung anlagenweiter Fehlerverursacher genutzt werden [72]. Das Verfahren ist jedoch lediglich für sinusförmig oszillierende Fehler, wie z.B. falsche Reglereinstellung, und lineare Teilprozesse anwendbar.

2.3.1 Grundlagen

Das Verfahren zur Online-Erkennung nutzt die Tatsache, dass im Falle eines Kreislaufprozesses, wie in Abbildung 2.1 gezeigt und einer hervorgerufenen anlagenweiten sinusförmigen Oszillation mit einer Frequenz f_1 , der Verursacher dieser Schwingung, die meiste Energie in dem Frequenzband um f_1 übertragen wird. Die in diesem Frequenzband übertragene Leistung lässt sich berechnen gemäß

$$e_{dist} = \int_{f_1 - \Delta f_k}^{f_1 + \Delta f_g} p_{yy} df, \quad (2.53)$$

wobei $\Delta f_k + \Delta f_g$ das Frequenzband um f_1 ist und P_{yy} die quadratischen Amplituden der Fouriertransformation einer Zeitreihe sind. Diese werden im Folgenden berechnet gemäß

$$p_{yy} = \frac{\mathbf{z} \cdot \bar{\mathbf{z}}}{n}, \quad (2.54)$$

wobei \mathbf{z} die komplexen Werte der FFT (*Fast Fourier Transformation*, Schnelle Fourier Transformation) und $\bar{\mathbf{z}}$ die konjugiert komplexen Werte sind. Um die übertragenen Leistungen im anlagenweiten Störungsfrequenzband aller Teilprozesse vergleichen zu können, empfiehlt es sich, diese zu normieren. Eine Möglichkeit dazu ist, die Störungsenergie auf die gesamte übertragene Energie e_{ges} zu beziehen. Diese berechnet sich unter Berücksichtigung der Nyquist-Frequenz (halbe Abtastfrequenz) $f_{Nyquist}$ gemäß

$$e_{ges} = \int_0^{f_{Nyquist}} p_{yy} df. \quad (2.55)$$

Daraus lässt sich die relativ übertragene Störungsenergie aufstellen als

$$\eta_f = \frac{e_{dist}}{e_{ges}}. \quad (2.56)$$

Ist eine genügend große anlagenweite Oszillation erkannt, so gilt jener Teilprozess als Verursacher, welcher den größten η_f Wert aufweist. Hierfür kann kein analytischer Beweis erbracht werden, jedoch ist dies in abgewandelter Form ein gebräuchliches

Verfahren zur Erkennung von Verursachern anlagenweiter Fehler, z.B. in [66]. Um die relative übertragene Störungsenergie für jeden Teilprozess berechnen zu können, müssen zunächst die Störungsfrequenz f_1 und das Frequenzband $\Delta f_k + \Delta f_g$ bestimmt werden [72]. Hierfür kann eine Monte-Carlo-Simulation genutzt werden, um signifikante p_{yy} -Werte im Spektrum eines skalierten Zeitverlaufs von insignifikanten Werten zu unterscheiden. Weisen alle Teilprozesse eine identische signifikante Frequenz auf, so ist diese als anlagenweite Störungsfrequenz anzusehen. Dies gilt lediglich, da a priori vorausgesetzt, wurde, dass alle Teilprozesse linear sind.

Die durchgeführte Monte-Carlo Simulation zeigt, dass für weißes Rauschen, welches per Definition keine signifikanten Amplituden besitzt, die quadratischen Amplituden im Spektrum exponentiell verteilt sind mit $\lambda = 1$ gemäß

$$p_{yy} \sim \exp(-\lambda) . \quad (2.57)$$

Die Güte dieser Aussage ist in Abbildung 2.25 beispielhaft gezeigt. Dort sind die relativen Häufigkeiten $f(p_{yy})$ von bestimmten quadratischen Amplituden p_{yy} von weißem Rauschen mit exponentiell verteilten Zufallszahlen ($\lambda = 1$) gegenübergestellt. Die Verläufe sind deckungsgleich.

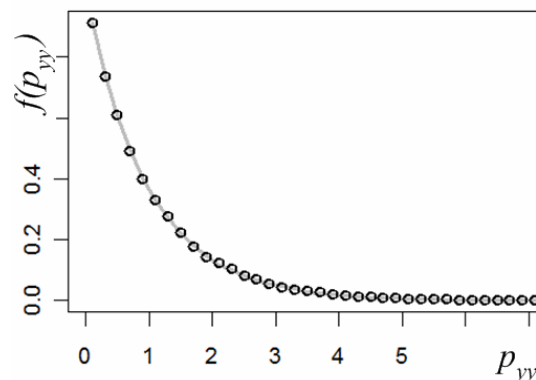


Abbildung 2.25: Amplitudenverteilung (durchgezogen: exponentiell verteilte Zufallszahlen, Punkte: Ergebnisse der Monte-Carlo-Simulation)

Unter der Kenntnis der zugrunde liegenden Verteilung und eines definierten Signifikanzniveaus α , lassen sich somit die Hypothesen bezüglich der Signifikanz einer einzelnen quadratischen Amplitude treffen.

- Nullhypothese: Für skalierte Zeitverläufe gilt somit unter der Kenntnis der Verteilung mit einer Unsicherheit von $\alpha = 0.005$, dass quadratische Amplituden, welche größer sind als 7.6, nicht von weißem Rauschen her stammen können und gelten somit als signifikant.
- Alternativhypothese: Quadratische Amplituden eines skalierten Zeitverlaufs, welche kleiner als 7.6 sind, können mit einer Wahrscheinlichkeit von $1 - \alpha$ durch weißes Rauschen bedingt sein. Sie gelten somit als insignifikant.

Eine Schwierigkeit dieses Online-Ansatzes ist, dass, bedingt durch den Kreislaufprozess, stets die Gefahr besteht, dass die Anlage in Resonanz gerät. Der Grund dafür liegt darin, dass der Anlageneintritt S_0 , welcher zur Verdeutlichung als ideal weißes Rauschen angenommen wird, eine Überlagerung vieler Frequenzen mit

einer sehr geringen Amplitude ist. Die Anlage wird somit auf vielen verschiedenen Frequenzen angeregt. Sämtliche Frequenzen, die jedoch nicht ein ganzzahliges Vielfaches dieser Resonanzfrequenz sind, werden eliminiert, da sie keine stehende Welle in der Anlage induzieren können. Das Frequenzspektrum einer jeden Anlage kann somit auch im normalen Betrieb signifikante Peaks (Spitzenwerte) im Spektrum aufweisen, hervorgerufen durch die Resonanzschwingung und deren Harmonischen (gerade oder ungerade Vielfache der Resonanzschwingung). Im Falle eines vorherrschenden oszillierenden anlagenweiten Fehlers tritt jedoch bei genügender Stärke des Fehlers die Tatsache auf, dass durch die vorangehende Normierung, die Amplituden der Resonanzfrequenz und deren Harmonischen unterhalb der Signifikanzgrenze fallen und somit nicht berücksichtigt werden müssen.

Bemerkung: Die Resonanzfrequenz einer Anlage und die Tatsache, ob gerade oder ungerade Harmonische auftreten, hängt von der Umlaufzeit T_{Circ} und der Art der Teilprozesse ab. Dies soll, aufgrund der Komplexität und der erforderlichen Ausführlichkeit jedoch nicht Teil dieser Arbeit sein.

2.3.2 Simulationsergebnisse

Um die Wirksamkeit der Methode zu zeigen, wird sie anhand des Benchmarkprozesses aus Abbildung 2.16 in einigen Simulationen getestet. In jeder Simulation variiert dabei der Verursacher. Auf den Ausgang des simulierten Verursachers wird eine Oszillation in der ersten Simulationsreihe mit der Frequenz 0.1 und der Amplitude 2 addiert. In der zweiten Simulationsreihe wird eine Oszillation mit der Frequenz 0.2 und der Amplitude 1 dazugaddiert. Das Resonanzspektrum der Messung von S_1 für den ungestörten Betrieb ist in Abbildung 2.26 dargestellt. Die Resonanzfrequenz beträgt bei dieser Anlage $T_{Circ} / 2 \approx 0.026$. Die Anlage zeigt aufgrund ihrer Struktur nur ungerade Harmonische. Die berechneten η_f -Werte bei einer Störungsfrequenz von 0.1 sind in Tabelle 2.2 gegeben und die η_f -Werte bei der Störungsfrequenz von 0.2 sind in Tabelle 2.3 zu finden.

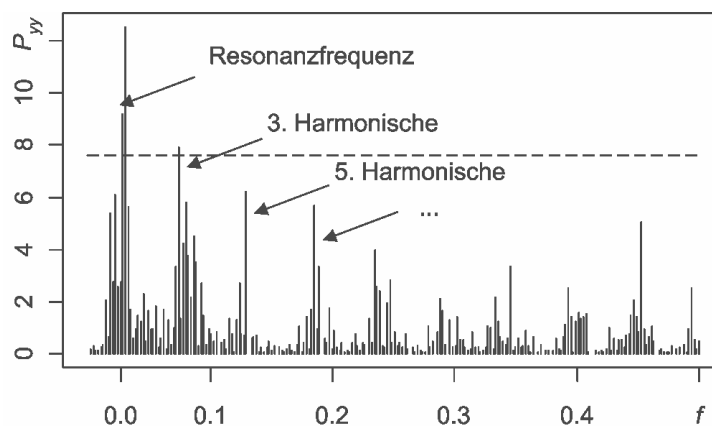


Abbildung 2.26: Resonanzspektrum S_1

Tabelle 2.2: Relative Störungsenergien bei der Störungsfrequenz 0.1 und Störungsamplitude 2

	Verursacher				
	S_1	S_2	S_3	S_4	S_5
$\eta_f(S_1)$	0.910	0.792	0.739	0.710	0.897
$\eta_f(S_2)$	0.888	0.907	0.710	0.705	0.873
$\eta_f(S_3)$	0.869	0.887	0.878	0.688	0.852
$\eta_f(S_4)$	0.860	0.872	0.860	0.859	0.840
$\eta_f(S_5)$	0.797	0.826	0.815	0.806	0.916

Tabelle 2.3: Relative Störungsenergien bei der Störungsfrequenz 0.2 und Störungsamplitude 1

	Verursacher				
	S_1	S_2	S_3	S_4	S_5
$\eta_f(S_1)$	0.696	0.489	0.482	0.457	0.694
$\eta_f(S_2)$	0.666	0.714	0.418	0.441	0.673
$\eta_f(S_3)$	0.645	0.671	0.665	0.387	0.651
$\eta_f(S_4)$	0.606	0.638	0.623	0.630	0.608
$\eta_f(S_5)$	0.514	0.547	0.547	0.552	0.734

Der Verursacher wird somit in jeder Simulation anhand seines hohen η_f -Wertes identifiziert.

2.3.3 Schätzung des Spektrums mit geringer Rechenleistung

Gemäß den industriellen Anforderungen aus Abschnitt 1.2 soll im Folgenden eine neue Alternative zum klassischen Spektrum mittels FFT gegeben werden. Die meisten Prozessleitsysteme haben keine fertigen Routinen zur Berechnung der FFT, besitzen aber in vielen Fällen einen Baustein zur FIR (*Finite Impulse Response*, Filter mit endlicher Impulsantwort). Unter Nutzung dieses Bausteins ist es möglich eine gute Annäherung eines Spektrums durch diese parallel geschalteten Bandpassfilter aufzustellen. Der schematische Ablauf ist in Abbildung 2.27 dargestellt.

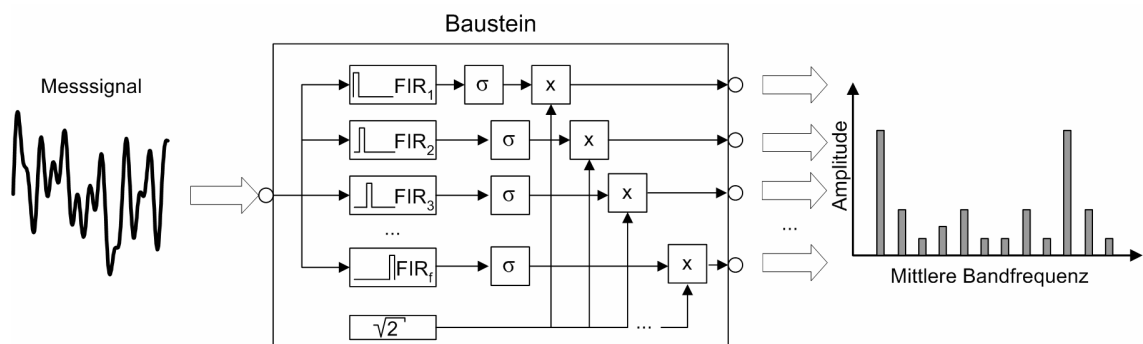


Abbildung 2.27: Schema zur Schätzung des Spektrums

Das gefilterte Signal eines jeden Bandpassfilters sollte im Idealfall nur noch aus einer Sinusschwingung bestehen, was zu folgenden Anforderungen an den Entwurf der Filter führt:

- Bandpassbreite so schmal wie möglich wählen,
- Filterordnung so groß wie möglich wählen.

Sind diese Anforderungen erfüllt, so lässt sich mittels des Zusammenhangs

$$a_{\sin} = \sqrt{2} \cdot \sigma_{\sin} \quad (2.58)$$

zwischen der Standardabweichung σ_{\sin} einer Sinusschwingung und der Amplitude a_{\sin} derselben Sinusschwingung eine gute Näherung des Spektrums schätzen. Sollte die direkte Möglichkeit zur Berechnung von FIR-Blöcken nicht bestehen, so lassen sich diese mittels eines rekursiven Butterworth-Filters selbst programmieren. Die Filterparameter können dabei in einer externen Datei gespeichert und bei jedem Aufruf neu geladen werden. In Abbildung 2.28 (oben) ist eine überlagerte Schwingung zu sehen, einmal mit dem klassischen FFT (unten links) Spektrum und einmal auf dem FIR basierten Verfahren (unten rechts) mit der Filterordnung 8 und der Bandbreite 0.01 (entspricht 50 äquidistanten Bandpassfiltern).

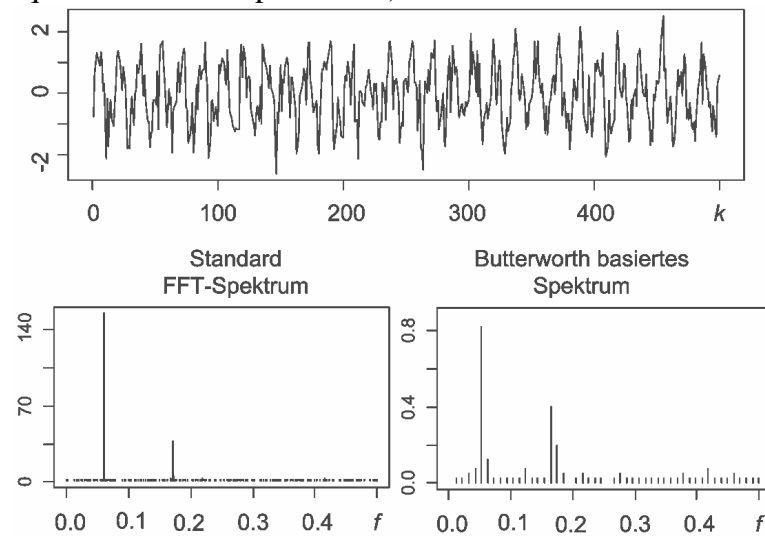


Abbildung 2.28: FFT-Spektrum (links) und geschätztes Spektrum (rechts, Filterordnung 8, Filterbreite 0.01)

Es ist zu sehen, dass beide Spektren Peaks bei denselben Frequenzen (~ 0.06 und ~ 0.17) aufweisen. Dennoch ist die Höhe der Peaks unterschiedlich. Dies ist dadurch bedingt, dass p_{yy} nicht ohne Umrechnung der eigentlichen Amplitude entspricht, wie sie im Butterworth basierten Spektrum zu sehen ist. Daher sind die Hypothesen und die daraus berechneten Grenzwerte nach 2.3.1 nicht anwendbar. Um dennoch Aussagen über die Grenzwerte zu ermöglichen und da es sich um eine nichtlineare Transformation zwischen p_{yy} und der Amplitude handelt, wurde die Monte-Carlo-Simulation aus Abbildung 2.25 noch einmal für das Butterworth basierte Spektrum durchgeführt. Da die Amplituden des Spektrums auch von der Breite des Butterworth-Filters abhängen, wurde die Anzahl an parallel geschalteten Filtern n_F ebenfalls variiert. In Abbildung 2.29 sind die kritischen Amplituden a_{crit} (Signifikanzniveau von 0.95) in Abhängig der Anzahl an parallel geschalteten Filter dargestellt.

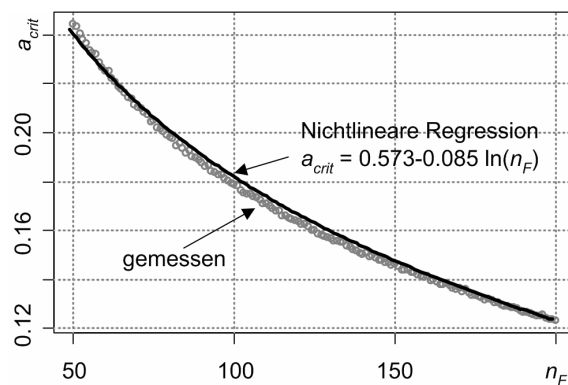


Abbildung 2.29: Kritische Amplituden zur Erkennung von signifikanten Amplituden auf Butterworth basiertem Spektrum

2.3.4 Online-Einsatz

Das genannte Verfahren ist für den industriellen Online-Einsatz programmiert worden und arbeitet nach den folgenden automatisierten Schritten:

1. Auslesen der aktuellen Prozessvariablen,
2. Suche nach anlagenweiter Oszillation und Identifikation der Frequenz f_1
3. Berechnung aller η_f -Werte
4. Bestimmung des Teilprozesses mit dem höchsten η_f -Wert

In Abbildung 2.30 ist die Benutzeroberfläche des entwickelten Programms gezeigt.

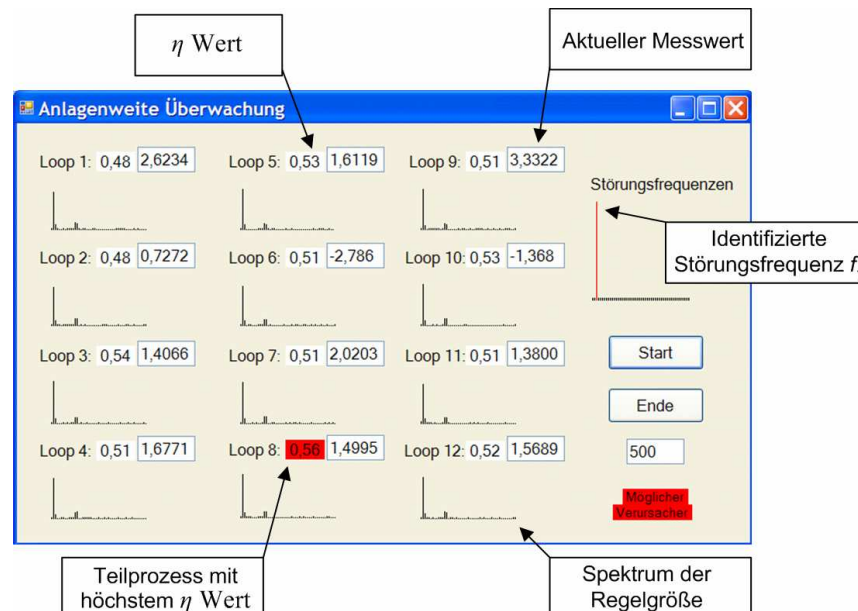


Abbildung 2.30: Benutzeroberfläche des Programms zur Verursachenerkennung

2.3.5 Laufzeitanalyse

Da für jeden Prozess lediglich ein Spektrum geschätzt werden muss und der η_f -Wert daraus bestimmt wird, gilt, dass die Rechenzeit (Laufzeit) des Online-Verfahrens linear von der Anzahl der Teilprozesse M abhängt. So gilt, dass

$$O(M) . \quad (2.59)$$

Außerdem gilt für die Laufzeitklasse, bezogen auf die Anzahl der Messungen, dass

$$O(n) . \quad (2.60)$$

Die Rechendauer für die Schätzung des Spektrums mit anschließender Berechnung der relativen Störungsenergie für einen Vektor der Länge n ist auf einem 1.8 GHz AMD Rechner mit 1 GB RAM bestimmt worden, gemäß Abbildung 2.31.

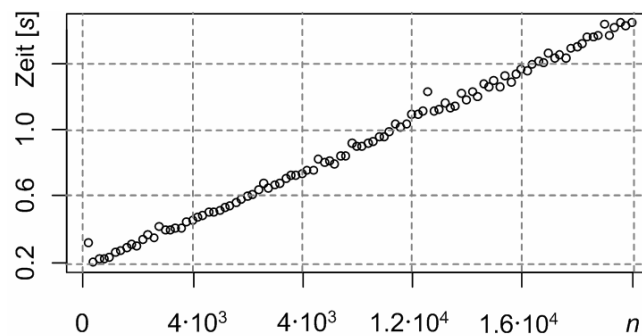


Abbildung 2.31: Rechenzeit des Online-Verfahrens

2.4 Zusammenfassende Bemerkung

In Kapitel 2 wurden zwei Verfahren zur Erkennung anlagenweiter Fehlerverursacher gezeigt. Der Vorteil dieser komplexen Herangehensweise liegt darin, dass nicht jeder Regelkreis diagnostiziert werden muss, sondern lediglich jener, der als Verursacher eines anlagenweiten Fehlers identifiziert wurde. Dies führt, wie beschrieben, zu schnelleren und verlässlicheren Ergebnissen. Außerdem kann die Kenntnis über den Verursacher ggf. zum Alarmmanagement genutzt werden. Anstatt den Anlagenfahrer auf alle Alarmer gleichzeitig reagieren zu lassen, muss er lediglich auf den Alarm des Verursachers reagieren. Der Nachteil dieses Ansatzes besteht (heutzutage) eindeutig in der Online-Anwendbarkeit, zumindest sofern es keinerlei vereinfachende Annahmen gibt. Kann davon ausgegangen werden, dass es eine messbare Totzeit zwischen allen Teilprozessen (Regelkreisen) der Anlage gibt, kann zumindest das völlig automatisierbare Offline-Verfahren aus Abschnitt 2.2.2 angewendet werden. Kann weiterhin davon ausgegangen werden, dass alle Prozesse linear sind, so ist bei genügend großer Abtastrate das Verfahren aus Abschnitt 2.2.1 anwendbar. Unter Annahme einer weiteren Einschränkung, dass es sich bei dem Fehler um eine hervorgerufene sinusförmige Oszillation handelt, so ist sogar das Online-Verfahren aus Abschnitt 2.3 anwendbar. Dieses besitzt marginale Ansprüche an Rechenleistung und Speicherplatz. Anhand von Benchmarkprozessen wurde gezeigt, dass alle Methoden, sofern die jeweiligen Anforderungen erfüllt sind, den Fehlerverursacher eindeutig identifizieren können.

Es sei jedoch erwähnt, dass der automatisierte Offline-Einsatz vorzuziehen ist, da bei der Bestimmung der Beeinflussungsrichtung auch sämtliche Totzeiten bestimmt werden. Diese können genutzt werden, um ein Fehlerfortpflanzungsmodell aufzustellen, welches Anlagenbetreibern hilft, ihr Prozessverständnis zu erhöhen und ggf. frühzeitig nach Fehlern intervenieren zu können, um deren Ausbreitung zu vermeiden. In

Abbildung 2.32 und Abbildung 2.33 sind die Fehlerfortpflanzungsmodelle der Matrizen \mathbf{Z}_1 (kein anlagenweiter Fehler) und \mathbf{Z}_2 (anlagenweiter Fehler mit Verursacher in S_j) aus Abschnitt 2.2 dargestellt.

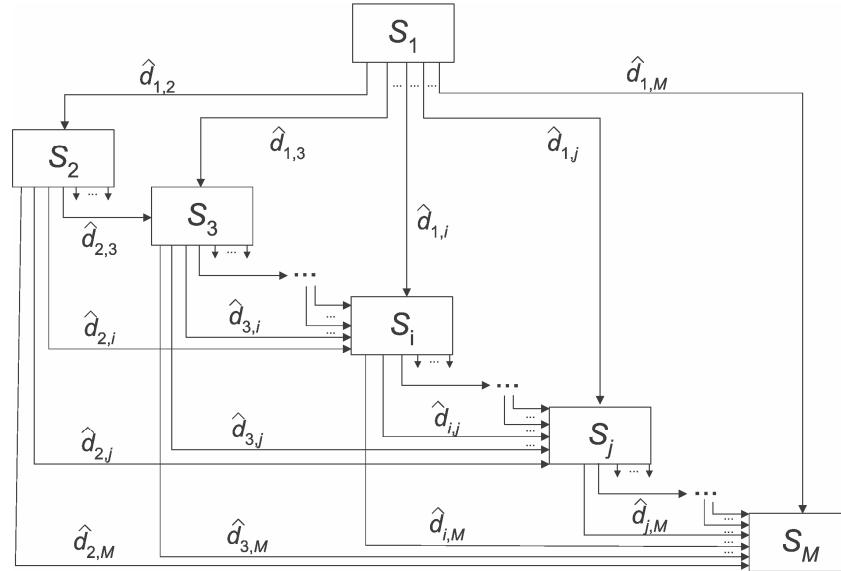


Abbildung 2.32: Fehlerfortpflanzungsmodell im Falle eines vorhandenen anlagenweiten Fehlers

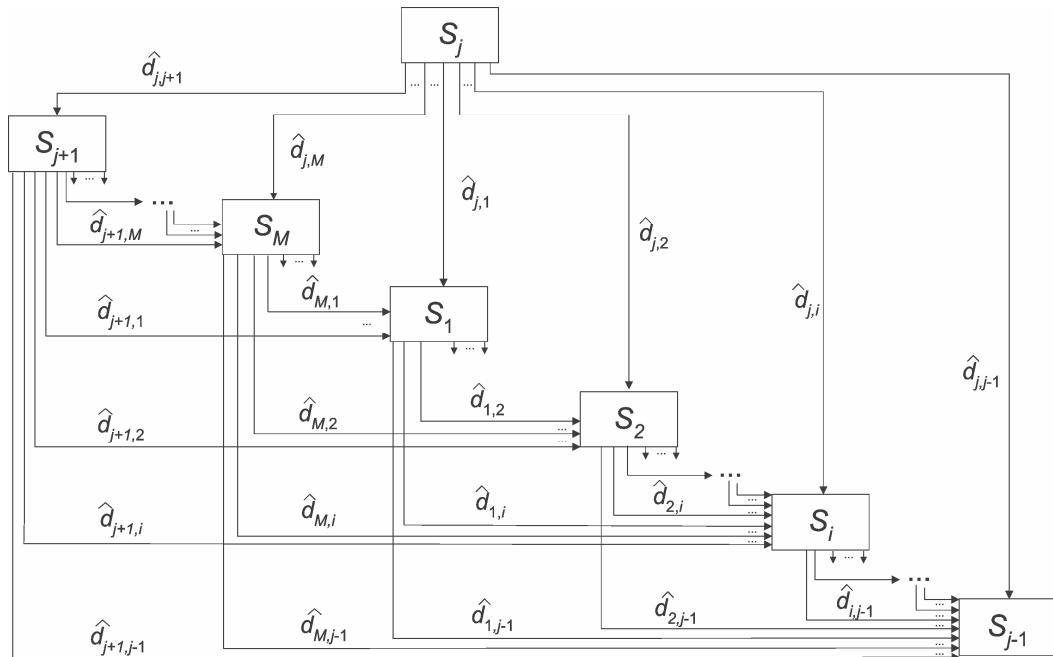


Abbildung 2.33: Fehlerfortpflanzungsmodell im Falle eines anlagenweiten Fehlers hervorgerufen durch S_j

Kapitel 3

Fehlererkennung einzelner Regelkreise

Basierend auf der in Abbildung 2.2 dargestellten Prozedur zur anlagenweiten Regelkreisdiagnostik folgt, dass, sofern es sich bei dem identifizierten Verursacher um den der externen Störung am nahe liegendsten Teilprozess S_1 handelt und die externe Störung S_0 bekannt ist, eine Fehlererkennung von S_0 durchgeführt werden muss. Unter der Annahme, dass es sich bei diesem Teilprozess S_0 um einen Regelkreis handelt, lassen sich die in Abschnitt 1.1 genannten Fehler noch einmal wie folgt genauer einteilen:

- Typ 1: Fehler, die das Potential haben einen anlagenweiten Fehler zu generieren. Hierfür sind hauptsächlich die Szenarien „Aggressive Regelung“, „Schwingender Sollwert“, „Haft- und Gleitreibung im Ventilkörper“ und „externe Störungen“ verantwortlich.
- Typ 2: Fehler, die in der Regel auf einen Regelkreis beschränkt sind, meistens hervorgerufen durch die Szenarien „Inaktive Regelungen“, „Beschränkungen“, „Messrauschen“ und „Langsame Regelungen“.

Bemerkung: In seltenen Fällen können jedoch auch Fehler des Typs 2 zu anlagenweiten Fehlern führen. So besteht z.B. bei inaktiven oder begrenzten Regelungen die Gefahr, dass auch die nachfolgenden Regelkreise ebenfalls an ihre Grenzen geführt werden und somit beschränkt arbeiten. Die Erkennung und Diagnose eines solchen anlagenweiten Szenarios soll jedoch nicht Teil der Arbeit sein, da es zum einen in der Industrie sehr selten auftritt und zum andern eine Verursacheridentifizierung mit den in Kapitel 2 genannten Verfahren unmöglich ist. Der Grund dafür ist, dass der beschränkte oder inaktive Regler die anderen Regelkreise nicht durch sein Eingreifen beeinflusst sondern durch seine Inaktivität. Um jedoch auch evtl. zukünftige Herangehensweise zur Verursacheridentifikation zu berücksichtigen, soll die Fehlererkennung für beide Fehlertypen ausgelegt sein.

Beide Fehlertypen zeichnen sich generell dadurch aus, dass die Regelung nicht optimal arbeitet, wodurch die einfachste Art der Erkennung somit in einer Einschätzung der einfachen Regelungsqualität besteht. Die bekanntesten Verfahren dazu sind in Abschnitt 3.1 aufgeführt. Der Nachteil dieser Ansätze besteht oft in den getroffenen Einschränkungen, der Komplexität und der Unfähigkeit zur völligen Automatisierung. Deshalb werden in Abschnitt 3.2 Verfahren zur speziellen Erkennung des Fehlertyps 1 gezeigt. Diese zeichnen sich, ausgenommen von bestimmten äußeren Störungen, immer durch eine hervorgerufene Oszillation aus, was deren Erkennung somit auf eine

Oszillationserkennung reduziert. Um die Wirksamkeit der einzelnen Ansätze zu zeigen, soll der in Abbildung 3.1 dargestellte Prozess genutzt werden. Die Abtastzeit beträgt 0.05.

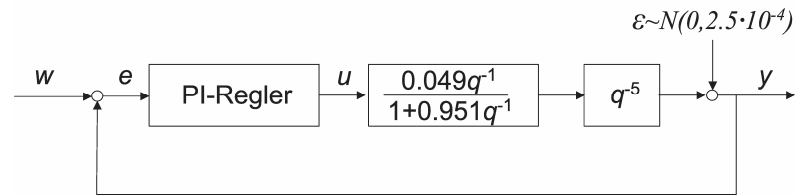


Abbildung 3.1: Prozess zur Wirksamkeitseinschätzung der Fehlererkennungsmethoden

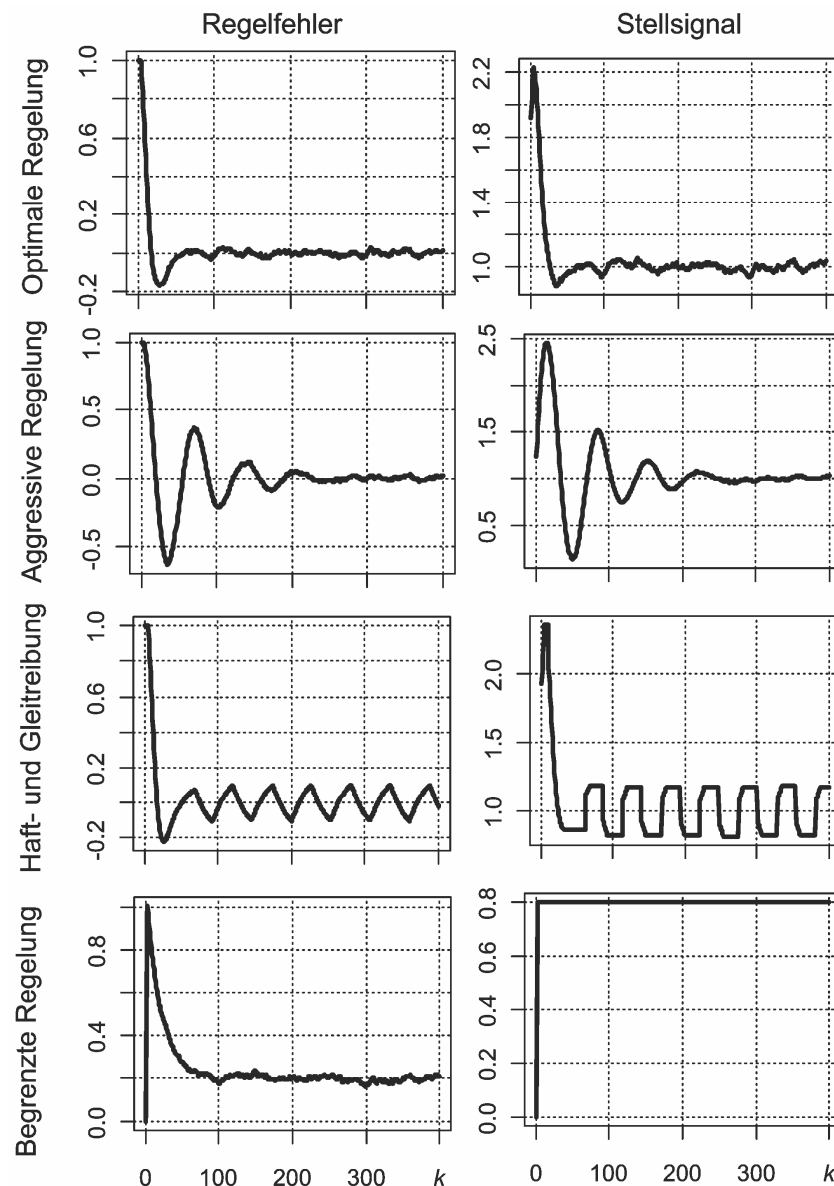


Abbildung 3.2: Verschiedene Regelungsszenarien nach Sollwertsprung von 0 auf 1

Für diesen Prozess sollen verschiedene Regelungsszenarien simuliert werden. Dazu sind in Abbildung 3.2 die Regelgrößenverläufe für einen nahezu optimal eingestellten PI-Regler und für die Regelungsszenarien „aggressive Regelung“, „Haft- und Gleitreibung“ und „begrenzte Regelung“ gezeigt. Der nahezu optimale Regler ist so

eingestellt, dass er eine maximale Überschwingung von 20% ($K_{PR} = 1.5$, $T_N = 0.7$) erreicht. Zur Realisierung der aggressiven Regelung wird ein Regler mit den Reglerparametern $K_{PR} = 0.6$ und $T_N = 0.19$ gewählt. Die Simulation des Ventils mit Haft- und Gleitreibung basiert auf der Methode von [5] mit den Parametern $S = 5$ und $J = 5$ (s. Abschnitt 4.3 für nähere Modellinformationen). Die beschränkte Regelung wird durch die Beschränkung des Stellsignals zwischen 0 und 0.8 erreicht.

3.1 Fehlererkennung mittels Performance-Indizes

Sowohl Fehler des Typs 1 als auch Fehler des Typs 2 zeichnen sich dadurch aus, dass die erreichte Qualität der Regelung deutlich unterhalb der erreichbaren Optimalqualität liegt. Die einfachste Erkennung der beiden Fehlertypen besteht somit darin, die Güte der Regelung zu beurteilen. Hierfür sollen drei Verfahren genannt werden.

3.1.1 Harris-Index

Der Harris-Index, oder auch Minimumvarianz-Kriterium genannt, ist in modifizierter Art Grundlage vieler kommerzieller Programme zur Bestimmung der Regelungsqualität, wie z.B. Loop ScoutTM von Honeywell, Process doctorTM der Firma Matrikon und Loop Optimizer SuiteTM von ABB. Der Index vergleicht die erreichte Regelgrößenvarianz eines Störgrößenreglers σ_y^2 für einen stochastischen Prozess mit der Varianz, die ein Benchmarkregler mit der theoretisch minimal möglichen Varianz σ_{MV}^2 , der so genannte Minimumvarianz-Regler (eingeführt durch [1]), erreicht hätte. Für die Berechnung des Index gilt, dass

$$\eta_{Harris} = \frac{\sigma_{MV}^2}{\sigma_y^2}. \quad (3.1)$$

Da beim Harris-Index lediglich das Störverhalten untersucht wird, kann der Sollwert mit Null angenommen werden. Die minimal erreichbare Varianz ergibt sich dann nach [63] gemäß

$$\sigma_{MV}^2 = (1 + f_1^2 + \dots + f_{d-1}^2) \sigma_\varepsilon^2, \quad (3.2)$$

wobei d der diskreten Totzeit entspricht, f_1 bis f_{d-1} die Impulsantwortkoeffizienten der Störübertragungsfunktion des Minimumvarianzreglers sind und σ_ε^2 die Varianz der stochastischen Störung ist. Zur Berechnung der minimalen Varianz müssen somit diese Parameter bekannt sein. Die Totzeit des geschlossenen Regelkreises kann entweder mittels der klassischen Verfahren (siehe z.B. [35]) oder aber mit dem in Abschnitt 2.2.2 genannten Verfahren zur nicht linearen Totzeitschätzung basierend auf KNNK bestimmt werden. Die Störung ε_k kann mithilfe eines *Whitening* Filters nach [23] durch einen ARMA-Filter gemäß

$$y_k = (-a_1 q^{-1} - \dots - a_q q^{-q}) y_k + (1 + \hat{f}_1 q^{-1} + \dots + \hat{f}_p q^{-p}) \hat{\varepsilon}_k \quad (3.3)$$

geschätzt werden. Die Filterordnungen können nach [3] gewählt werden oder durch Versuch und Irrtum bestimmt werden. Ist $\hat{\varepsilon}_k$ geschätzt, so lässt sich η_{Harris} im letzten Schritt nach [23] berechnen gemäß

$$\eta_{Harris} = cc f_{\varepsilon, y}^{*2}(0) + cc f_{\varepsilon, y}^{*2}(1) + \dots + cc f_{\varepsilon, y}^{*2}(d-1), \quad (3.4)$$

wobei $cc f_{\varepsilon, y}^{*2}(i)$ dem normierten Kreuzkorrelationswert zwischen ε und y bei der diskreten Verschiebung $\kappa = i$ entspricht. Die genaue Vorgehensweise kann in [49] für verschiedene Beispiele gefunden werden und soll exemplarisch für einen PI-geregelten ($K_{PR} = 2.25$, $T_N = 1.33$) Prozess erster Ordnung mit Totzeit $\left(G_p(s) = \frac{1}{0.75s+1} \cdot e^{-s0.15}, \Delta T = 0.05 \right)$ gezeigt werden. Die skalierten Daten sind in Abbildung 3.3 dargestellt.

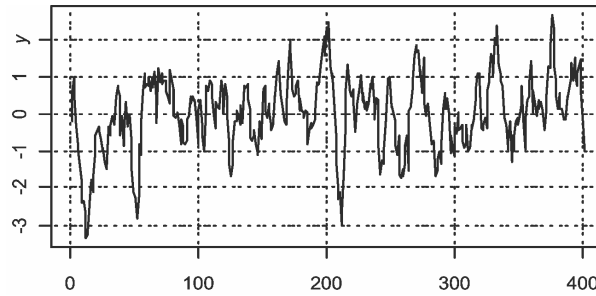


Abbildung 3.3: Regelgrößenverlauf bei gut eingestelltem Regler

Es lässt sich ein ARMA-Modell zweiter Ordnung für den Prozess schätzen gemäß

$$y_k = \frac{1 - 0.7420q^{-1} + 0.0805q^{-2}}{1 - 1.6440q^{-1} + 0.7155q^{-2}} \varepsilon_k. \quad (3.5)$$

Für $d = 3$ ergibt sich, basierend auf den berechneten Kreuzkorrelationswerten, der Harris-Index gemäß

$$\eta_{Harris} = 0.5047^2 + 0.4544^2 + 0.4058^2 = 0.6258. \quad (3.6)$$

Ein Wert η_{Harris} von 1 entspricht somit der Performance des Minimumvarianzreglers und ein Wert von nahe 0 somit einer schlechten Regelung (im Sinne des Kriteriums).

Die Verwendung des Harris-Index ist nur begrenzt in dieser allgemeinen Form zur Performance-Abschätzung geeignet, so lassen sich aus [59], [58] und [49] folgende Einschränkungen zusammenfassen:

1. Regelmäßige Oszillationen der Regelgröße können zu fehlerhaften Einschätzungen führen.
2. PI(D)-Regler sind nicht in der Lage η_{Harris} von 1 zu erreichen.
3. Minimumvarianzregler für höhere Prozessordnungen oder große Totzeiten sind nicht realisierbar.
4. Eine Optimierung hin zum Minimumvarianzregler führt stets zu einer erhöhten Stellgeräteaktivität und somit zu einem höheren Verschleiß ([63] erweitert die

Grundidee des Minimumvarianzreglers um eine Gewichtung der Stellgerätearbeit).

5. Minimumvarianzregler sind nicht robust im Bezug auf zeitvariante Prozesse ([33] zeigt ein Verfahren welches in diesem Fall anzuwenden ist).
6. Hohe Datenkompression führt zu fehlerhaften Einschätzungen.
7. Die stochastischen Störungen müssen stationär sein.

Der Bezug auf die Performance des Minimumvarianzreglers ist somit in vielen Fällen nicht geeignet. Als besseres Verfahren schlägt [41] beispielsweise den Bezug auf einen optimalen PI-Regler vor und [19] den Bezug auf einen vergleichbaren (als gut identifizierten) Benchmarkprozess. Der Vergleich mit einem Benchmarkprozess wird in Abschnitt 3.1.3 noch genauer erläutert.

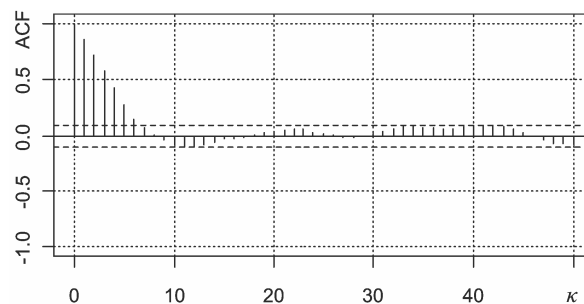


Abbildung 3.4: Autokorrelationsfunktion der Regelgröße bei gut eingestelltem Regler mit der Totzeit 5

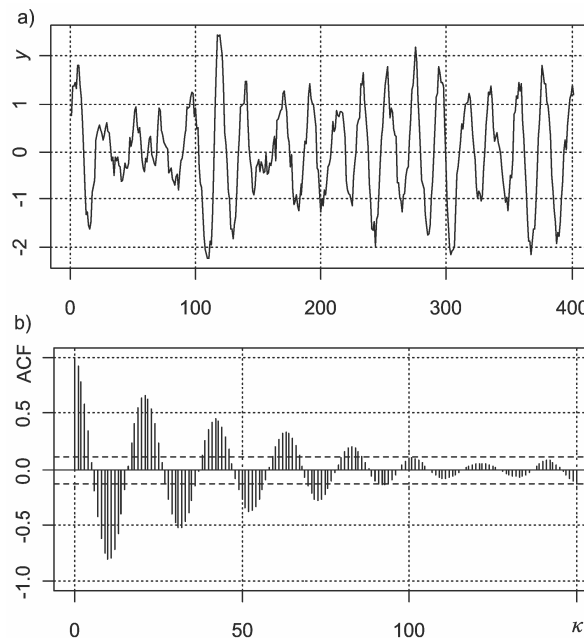


Abbildung 3.5: a) Skalierter Regelgrößenverlauf und b) Autokorrelationsfunktion bei aggressiv eingestelltem Regler

Neben den genannten Einschränkungen ist die Berechnung des Harris-Index ebenfalls nur mithilfe spezieller Software möglich, daher schlagen sowohl [19] als auch [49] die Auswertung der Autokorrelationsfunktion vor. Enthält diese signifikante Werte für $\kappa > d$, so kann von einer schlechten Regelung ausgegangen werden. Nach [3] gelten Autokorrelationswerte als insignifikant, wenn sie bei n Messwerten kleiner als

$2/\sqrt{n}$ sind. Für den in Abbildung 3.3 dargestellten Prozess ergibt sich die Autokorrelationsfunktion aus Abbildung 3.4 mit der strichierten Signifikanzgrenze. Bereits bei der Verschiebung $\kappa=7$ ist der berechnete Autokorrelationswert insignifikant.

Wird der Regler für denselben Prozess aggressiver eingestellt ($K_{PR} = 2.5$, $T_N = 0.45$), so ergibt sich der skalierte Regelgrößenverlauf aus Abbildung 3.5 a). Der Harris-Index berechnet sich hierfür zu $\eta_{Harris} = 0.2673$. Die berechnete Autokorrelationsfunktion ist in Abbildung 3.5 b) dargestellt. In diesem Fall wird erst bei einer Verschiebung von $\kappa=83$ ein insignifikanter Autokorrelationswert erreicht.

Für die simulierten Regelungsszenarien aus Abbildung 3.2 ergeben sich die Autokorrelationsfunktionen aus Abbildung 3.6 mit den dort dargestellten Harris-Indizes.

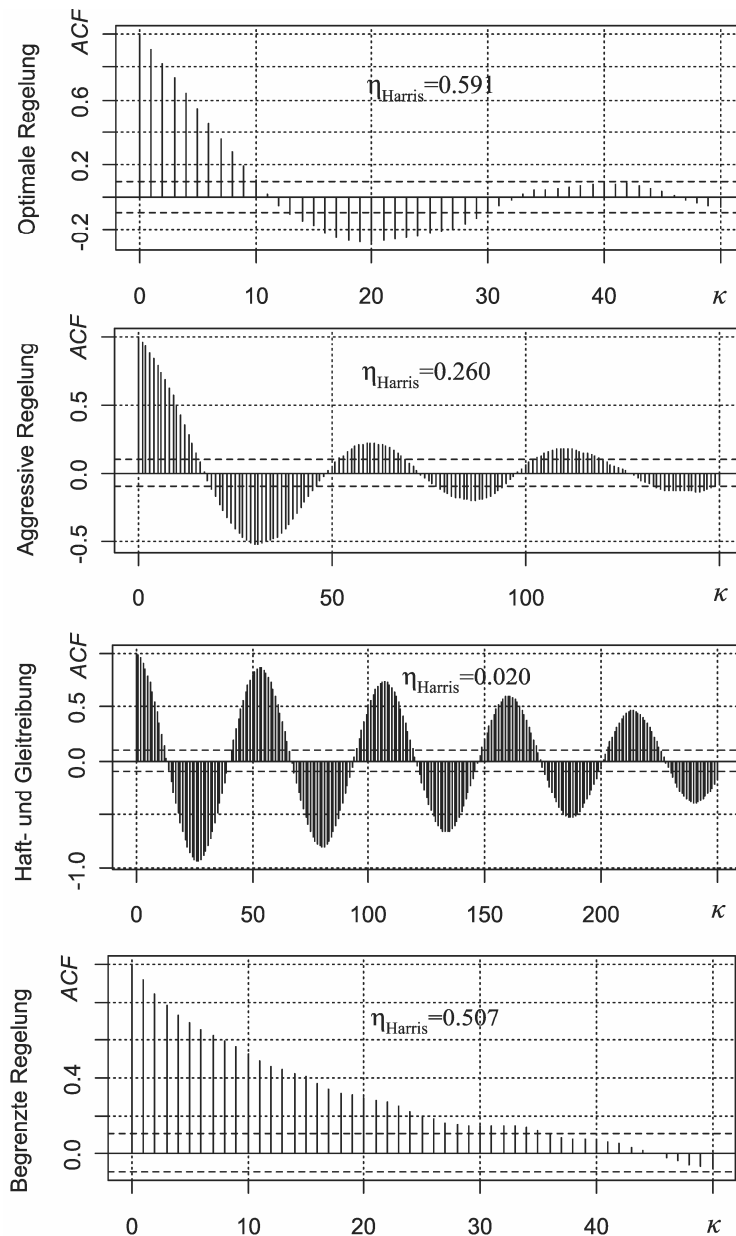


Abbildung 3.6: Autokorrelationsverläufe der Regelgrößen für die verschiedenen Regelungsszenarien mit berechneten Harris-Indizes

3.1.2 Vorhersage des Regelfehlers

Eine zweite Möglichkeit die einfache Regelgüte zu bestimmen liefert der *CLPA*-Index (*Control Loop Performance Assessment*, Regelkreis Performance Abschätzung) nach [11] und [59]. Dieser Index basiert auf der ähnlichen Überlegung wie der Harris-Index, wobei angenommen wird, dass im Falle einer idealen Regelung, der Regelfehler e_k nach einer Ausregelung nicht über einen gewissen Horizont b (Totzeit zuzüglich Ausregelzeit) hinaus vorhersagbar ist. Die Schätzung des zukünftigen Regelfehlers \hat{e}_{k+b} wird somit im Falle einer guten Regelung stark vom tatsächlichen Regelfehler abweichen. Um den zukünftigen Regelfehler zu schätzen, wird in [11] vorgeschlagen, den AR (autoregressiver)-Ansatz p -ter Ordnung gemäß

$$\hat{e}(k+b) = a_0 + a_1 e_k + a_2 e_{k-1} + a_3 e_{k-2} + \dots + a_p e_{k-p} \quad (3.7)$$

zu wählen, wobei a_0, a_1, \dots, a_p geschätzte Parameter sind. Für den Vergleich zwischen tatsächlichem und geschätztem Regelfehler wird ferner die Variable r_k eingeführt gemäß

$$r_k = e_k - \hat{e}_k. \quad (3.8)$$

Der CLPA Index η_{CLPA} berechnet sich direkt aus r_k gemäß

$$\eta_{CLPA} = 1 - \frac{\sigma^2(r)}{MSE(e)}. \quad (3.9)$$

Eine gute Regelung zeichnet sich somit nach (3.9) durch einen η_{CLPA} Wert nahe eins aus. Für die in Abbildung 3.2 gezeigten Szenarien ergeben sich die folgenden CLPA Indizes: Optimaler Regler: $\eta_{CLPA} = 0$, aggressiver Regler: $\eta_{CLPA} = 0.866$, Ventil mit Haft- und Gleitreibung: $\eta_{CLPA} = 0.224$, begrenzte Regelung: $\eta_{CLPA} = 0$. Die Schätzungen der Regelgrößenverläufe während der Ausregelung sind in Abbildung 3.7 dargestellt.

Bei der Betrachtung der berechneten CLPA-Indizes fällt auf, dass die optimale Regelung und die aggressive Regelung den Erwartungen entsprechen und somit korrekt eingeschätzt werden können. Die begrenzte Regelung unterscheidet sich, bezogen auf die Vorhersagbarkeit des Regelfehlers, jedoch nicht von der optimalen Regelung. Der Grund liegt darin, dass auch hierbei die Streuung und der Verlauf des zukünftigen Regelfehlers lediglich durch das aktuelle und zukünftige Messrauschen beeinflusst wird. Bei der aggressiven Regelung hingegen wird die Streuung des zukünftigen Regelfehlers hauptsächlich durch die Einstellung des Reglers bestimmt und ist somit in gewissen Grenzen vorhersagbar. Die Regelung mit einem simulierten Ventil mit Haft- und Gleitreibung entspricht - bei alleiniger Betrachtung des CLPA-Indizes - ebenfalls eher einer guten Regelung. Der Grund hierfür liegt darin, dass die Streuung und der Verlauf des Regelfehlers zwar eindeutig vorhersagbar sind, nicht jedoch durch ein einfaches AR-Polynom. Der Grund hierfür ist in der Unstetigkeit innerhalb des simulierten Ventils zu finden. Abschließend lässt sich über diese Art der Fehlererkennung sagen, dass sie nur für wenige Regelungsszenarien anwendbar ist. Dazu gehören sicherlich die aggressive Regelung, ein harmonisch schwingender Sollwert und harmonisch schwingende externe Störungen. Bei allen anderen Fehlern ist

dieses Verfahren nicht generell anwendbar. Doch selbst die Erkennung jener detektierbaren Regelungsszenarien ist nur unter Kenntnis der Einstellparameter b und p möglich. Für verschiedene Bereiche der Petrochemie können diese [59] entnommen werden. Als automatisierte Fehlererkennung ist dieses Verfahren jedoch zusammenfassend nicht verwendbar.

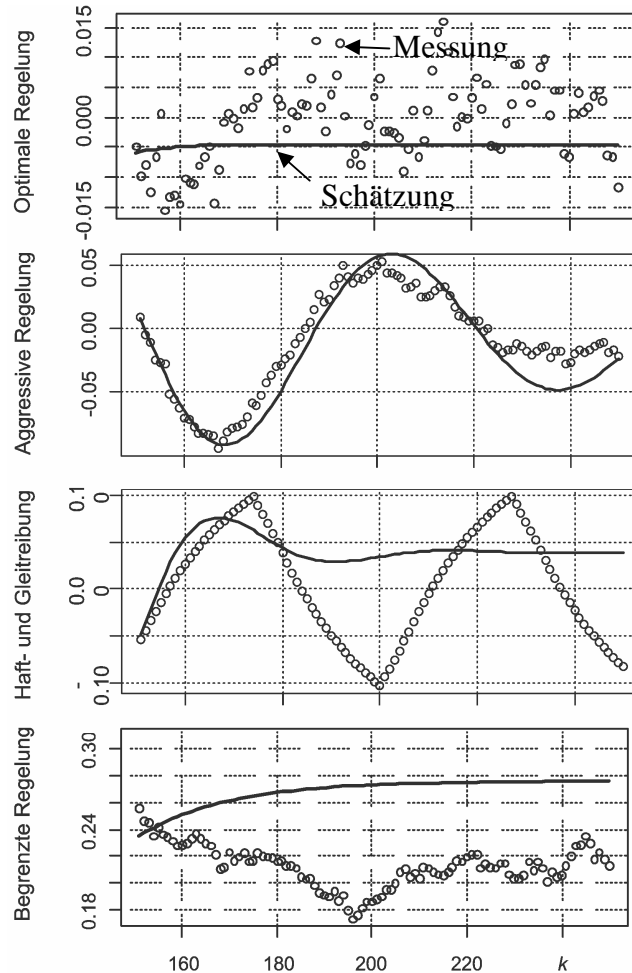


Abbildung 3.7: Verschiedene Regelungsszenarien zum Testen der Fehlererkennung

3.1.3 Berechnung individueller Maßzahlen

Eine weitere Möglichkeit zur Fehlererkennung wurde in ähnlicher Art in Abschnitt 3.1.1 als Verbesserungsvorschlag zum Harris-Index genannt. Es handelt sich dabei um den direkten Vergleich mit (vorher als gut definierten) Benchmarkprozessen. Dies ist unter anderem möglich durch die Berechnung individueller Maßzahlen, etwa wie in [61] angegeben als

$$c_{lp} = w_1 \cdot \sigma_e^2 + w_2 \cdot t_C + w_3 \cdot \int_0^t |\Delta u| d\tau \cdot t^{-1} + w_4 \cdot f_p, \quad (3.10)$$

wobei w_1 bis w_4 Gewichtungsfaktoren sind, σ_e^2 die Regelfehlervarianz ist, t_C die Ausregelzeit (z.B. stationäres Erreichen von 95% des Sollwerts) ist, $\int_0^t |\Delta u| d\tau \cdot t^{-1}$ die auf

einen Zeitraum normierte Stellgerätearbeit angibt und f_p der Fehlerfortpflanzung des zu untersuchenden Regelkreises entspricht (genauere Beschreibungen werden im Original nicht genannt).

Der Einsatz solcher Maßzahlen ist in der Industrie weit verbreitet, erfordert aber ein hohes Maß an Erfahrung und das Vorhandensein eines vergleichbaren Benchmarks.

Für die in Abbildung 3.2 gezeigten Prozesse soll dazu zunächst der c_{lp} Index für den optimalen Regler bestimmt werden. Da die Fehlerfortpflanzung in diesem einfachen Fall nicht berücksichtigt werden soll, wird der Gewichtungsfaktor w_4 zu Null gesetzt.

Die berechneten σ_e^2 , $\frac{1}{t} \int_0^t |\Delta u(\tau)| d\tau$ und t_C Werte ergeben sich für den optimalen Benchmarkprozess zu

$$\begin{aligned}\sigma_e^2 &= 0.022 \\ t_C &= 2.65 \\ \frac{1}{t} \int_0^t |\Delta u(\tau)| d\tau &= 0.259.\end{aligned}\tag{3.11}$$

Um alle berechneten Größen ungefähr gleich zu gewichten und eine Normierung zu erreichen, werden die folgenden Gewichtungsfaktoren verwendet: $w_1 = 15.37$, $w_2 = 0.13$ und $w_3 = 1.29$ ($w_4 = 0$). Damit ergibt sich für den optimalen Benchmarkprozess ein c_{lp} -Wert von 1.01. Die berechneten Werte der simulierten Regelungsszenarien sind Tabelle 3.1 zu entnehmen. Sowohl das Regelungsszenario „Haft- und Gleitreibung“ als auch das Szenario „Begrenzungen“ besitzen aufgrund ihrer Eigenarten eine unendlich große Ausregelzeit.

Tabelle 3.1: c_{lp} -Werte der verschiedenen Regelungsszenarien

	Optimale Regelung	Aggressive Regelung	Haft- und Gleitreibung	Begrenzte Regelung
σ_e^2	0.022	0.045	0.027	0.020
$\frac{1}{t} \int_0^t \Delta u(\tau) d\tau$	0.259	0.348	0.393	0.032
t_C	2.65	10.75	∞	∞
c_{lp}	1.01	2.53	∞	∞

Sofern es Maßstäbe bezüglich der Güte gibt, z.B. durch vergleichbare optimale Benchmarkprozesse, bietet die Berechnung individueller Maßzahlen eine gute Möglichkeit zur Fehlererkennung beider Fehlerarten. In diesem Fall ist die Berechnung ebenfalls vollständig automatisierbar.

3.2 Fehlererkennung mittels Oszillationserkennung

Die Grenzen und Einschränkungen der Fehlererkennung mittels Performance-Indizes zur Einschätzung der Regelungsgüte wurden in Abschnitt 3.1 gezeigt. Als bestes Verfahren hat sich dabei die Aufstellung individueller Maßzahlen erwiesen. Diese erfordern jedoch die Kenntnis eines optimalen Benchmarkprozesses. Ist dieser nicht

bekannt, lässt sich keine automatisierbare und verlässliche Fehlererkennung realisieren. Der vorliegende Abschnitt nutzt bei der automatisierten Fehlererkennung die Tatsache, dass die meisten anlagenweiten Fehler [66] sich durch das Charakteristikum „Oszillation“ auszeichnen. Die Fehlererkennung ist somit in den meisten Fällen auf eine Oszillationserkennung zurückzuführen. Wichtig hierbei ist, dass diese unabhängig von der Form der Schwingung ist, da verschiedene Szenarien auch verschiedene Schwingungsformen hervorrufen. Die Problematik einer Oszillationserkennung besteht immer darin, dass Messrauschen die Ergebnisse stark beeinflusst und die Entscheidung somit verfälschen kann. Dies soll für das Verfahren von [57] gezeigt werden, welches die Regelmäßigkeit der Nulldurchgänge eines mittelwertfreien Signals zur Oszillationserkennung nutzt. Der dafür entwickelte Index q_0 berechnet sich gemäß

$$q_0 = \frac{1}{3} \cdot \frac{\overline{\Delta k_0}}{\sigma_{k_0}}, \quad (3.12)$$

wobei $\overline{\Delta k_0}$ dem Mittelwert und σ_{k_0} der Standardabweichung der Zeit zwischen zwei Nulldurchgängen entspricht. Für das verrauschte, sinusförmige Messsignal x aus Abbildung 3.8 ergibt sich der q_0 -Wert gemäß $q_0 = \frac{1}{3} \cdot \frac{9.92}{15.15} = 0.218$.

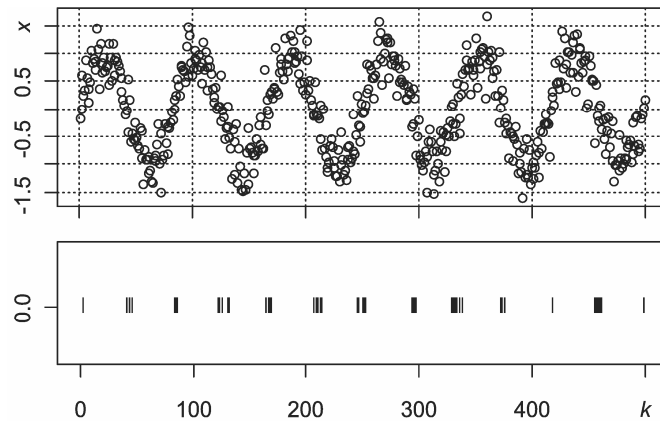


Abbildung 3.8: Signal (oben) mit dargestellten Nulldurchgängen (unten)

Nach [57] wird ein Signal vom menschlichen Gehirn als schwingend erkannt, sofern der q_0 -Wert größer als 1.3 ist. Das Signal aus Abbildung 3.8 würde somit bedingt durch das Messrauschen als „nicht oszillierend“ klassifiziert werden. Um das Problem zu umgehen, soll nach [57] zuvor die Autokorrelation des Signals berechnet werden. Der Grund hierfür liegt darin, dass die Autokorrelationsfunktion eines harmonisch schwingenden, mittelwertfreien, verrauschten Signals dieselbe Schwingungsdauer besitzt wie die Schwingung im Zeitbereich, weißes Rauschen jedoch weitgehend herausgefiltert wird. Für das verrauschte Signal aus Abbildung 3.8 ergibt sich der Autokorrelationsverlauf und den erkannten Nulldurchgängen in Abbildung 3.9. Basierend auf der Autokorrelation ergibt sich ein Oszillationsindex von $q_0 = \frac{1}{3} \cdot \frac{41.667}{0.516} = 26.89$. Das Abklingen des Autokorrelationsverlaufs ist durch die begrenzte Anzahl an Messdaten zu erklären, was jedoch die Periodizitätserkennung nicht beeinflusst.

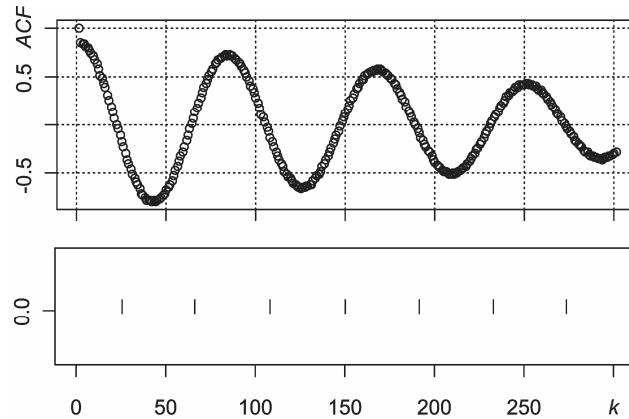


Abbildung 3.9: Autokorrelationsfunktion eines Signals (oben) mit dargestellten Nulldurchgängen (unten)

Der einzige Nachteil dieses Verfahrens ist, dass der Wert des Oszillationsindex nicht begrenzt ist, was eine automatische Auswertung oder eine theoretische Validierung erschwert. Daher wird im Folgenden der normalisierte Oszillationsindex n_{oi} [71] eingeführt. Er berechnet sich für den i -ten Nulldurchgang gemäß

$$n_{oi,i} = \begin{cases} \Delta k_{0,i} / \overline{\Delta k_0}, & \text{wenn } \Delta k_{0,i} \leq \overline{\Delta k_0} \\ \overline{\Delta k_0} / \Delta k_{0,i}, & \text{sonst} \end{cases}, \quad (3.13)$$

wobei $\Delta k_{0,i}$ die Anzahl an Abtastpunkten zwischen dem i -ten und dem $(i+1)$ -ten Nulldurchgang ist. Der normalisierte Oszillationsindex lässt sich als Mittelwert aller $n_{oi,i}$ -Werte

bestimmen gemäß

$$n_{oi} = \frac{1}{l-1} \sum_{i=1}^{l-1} n_{oi,i}, \quad (3.14)$$

wobei l der Anzahl an Nulldurchgängen entspricht. Liegen die berechneten n_{oi} -Werte über 0.71, so weisen diese auf eine signifikante Periodizität hin, wie sie nicht von weißem Rauschen hätte entstehen können. Die Herleitung dieses Grenzwerts soll im Folgenden gezeigt werden.

Da der Grenzwert nicht analytisch hergeleitet werden kann, soll eine Monte-Carlo Simulation genutzt werden. Dazu werden 500 Zeitreihen generiert, die jeweils aus 5000 Realisierungen der Zufallsvariablen X bestehen. Anschließend wird der zugehörige n_{oi} -Wert berechnet. Die Verteilung der 500 n_{oi} -Werte ist in Abbildung 3.10 dargestellt.

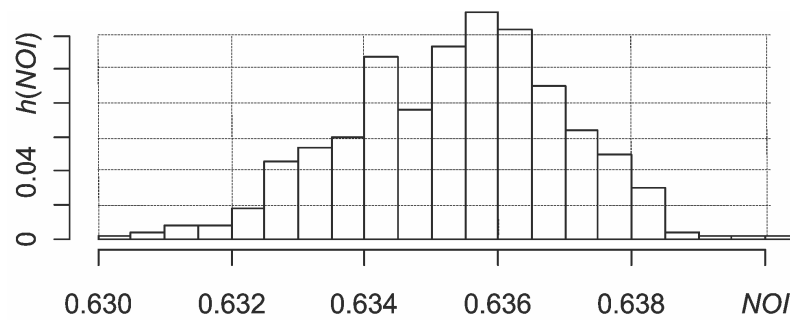


Abbildung 3.10: Relative Häufigkeiten als Ergebnis der Monte-Carlo Simulation bei 5000 Realisierungen

Basierend auf dieser empirisch ermittelten Verteilung lässt sich ein kritischer Wert $n_{oi,krit,5000} = 0.639$ ablesen, der größer ist als 99.5% der Daten. Es ist offensichtlich, dass bedingt durch das Gesetz der großen Zahlen die Streuung der n_{oi} -Werte für eine geringere Anzahl an Realisierungen (hier 5000) zunimmt. Werden die n_{oi} -Werte lediglich anhand von 100 Realisierungen bestimmt, so ergibt sich die Verteilung aus Abbildung 3.11 und der kritische Wert lässt sich bestimmen als $n_{oi,krit,100} = 0.704$. Da eine Berechnung des n_{oi} -Index für weniger als 100 Messwerte nicht empfehlenswert ist, da es in aller Regel zu wenige Nulldurchgänge oder die mittlere Zeit zwischen zwei Nulldurchgängen zu klein wird, soll als kritischer Wert $n_{oi,krit} = 0.71$ herangezogen werden. n_{oi} -Werte größer als 0.71 können somit nicht durch die stochastische Periodizität von weißem Rauschen hervorgerufen werden und gelten somit als Anzeichen für eine deterministische Periodizität.

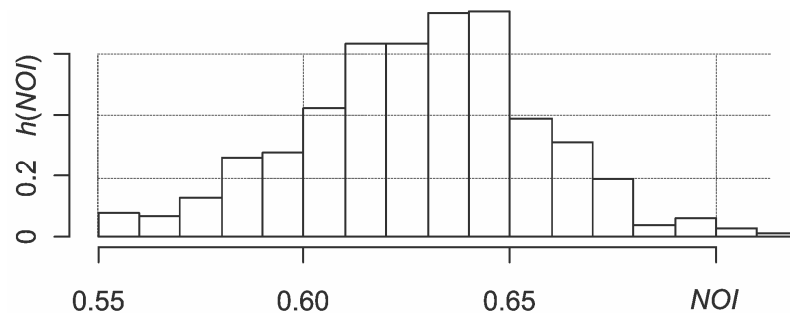


Abbildung 3.11: Relative Häufigkeiten als Ergebnis der Monte-Carlo Simulation bei 100 Realisierungen

Auch die Berechnung des normalisierten Oszillationsindex sollte auf den Werten der Autokorrelationsfunktion basieren. Für den Verlauf aus Abbildung 3.9 ergibt sich ein n_{oi} -Wert von 0.98. Für die verschiedenen Regelungsszenarien aus Abbildung 3.2 werden die n_{oi} -Werte basierend auf der Autokorrelationsfunktion des Regelfehlers berechnet. Die Autokorrelationsfunktion wird dazu bis zu einer Verschiebung von $\kappa_{\max} = 300$ berechnet. Die bestimmten Werte sind Tabelle 3.2 zu entnehmen. Die begrenzte Regelung weist lediglich drei Nulldurchgänge ($\kappa_{0,1} = 79$, $\kappa_{0,2} = 101$ und $\kappa_{0,3} = 147$) auf. Der n_{oi} -Wert kann für diesen Fall nicht verlässlich bestimmt werden.

Tabelle 3.2: n_{oi} -Werte der verschiedenen Regelungsszenarien

	Optimale Regelung	Aggressive Regelung	Haft- und Gleitreibung	Begrenzte Regelung
n_{oi}	0.579	0.766	0.981	[]

Die Fehlererkennung basierend auf der Oszillationserkennung ist, ebenso wie der Harris-Index und der *CLPA*-Index, nicht in der Lage die begrenzte Regelung als Fehler zu erkennen. Im Gegensatz zur Erkennung basierend auf dem *CLPA*-Index, ist das oszillationserkennende Verfahren auch in der Lage, nicht harmonische Schwingungen zu detektieren, wie sie etwa durch Unstetigkeiten bei der Haft- und Gleitreibung vorkommen. Die optimale Regelung lässt sich, ebenso wie bei der *CLPA*-Berechnung, gut identifizieren. Im Gegensatz dazu liefert der Harris-Index für den gut eingestellten Regler einen Wert, der weit unterhalb von 1 liegt und somit nicht auf eine gute Regelung hinweist, zumindest im Sinne des Minimumvarianz-Kriteriums.

Kapitel 4

Allgemeine Fehlerdiagnose einzelner Regelkreise

Basierend auf der in Abbildung 2.2 dargestellten Prozedur zur anlagenweiten Regelkreisdiagnostik steht am Ende jeder Auswertung eine ausführliche Fehlerdiagnose des betreffenden Teilprozesses. Unter der Annahme, dass es sich bei diesem Teilprozess um einen Regelkreis handelt, lässt sich diese Fehlerdiagnose im ersten Schritt als Unterscheidung der verschiedenen Charakteristika ansehen. Im zweiten Schritt, der Klassifikation, wird die Tatsache genutzt, dass jedes der in Abschnitt 1.1 genannten Regelungsszenarien ein typisches Spektrum an Charakteristika zeigt. Somit ist es möglich anhand der erkannten Charakteristika auf die einzelnen Szenarien zu schließen. Diese sind folgend noch einmal aufgezählt:

- aggressive Reglereinstellung,
- schwingender Sollwert,
- Ventilprobleme,
- externe Störungen,
- Messrauschen,
- Begrenzungen,
- inaktive Regelung,
- träge Reglereinstellung.

Wie bereits in Kapitel 3 erwähnt, können unter gewissen Umständen alle Szenarien zu anlagenweiten Fehlern führen. Allerdings ist es nicht möglich den Verursacher der Szenarien „Begrenzungen“, „Inaktive Regelung“ und „Träge Reglereinstellung“ mit den in Kapitel 2 genannten Verfahren zu identifizieren.

4.1 Problemstellung

Der vorliegende Abschnitt beschäftigt sich mit der ausführlichen Fehlerdiagnose einzelner Regelkreise unter Berücksichtigung der in Abschnitt 1.2 genannten Anforderungen. Dazu werden zunächst in Abschnitt 4.2 die gängigsten Regelungsszenarien beschrieben und deren Charakteristika aufgezählt. In Abschnitt 4.3 wird das Szenario „Haft- und Gleitreibung“ genauer mit seinen Charakteristika beschrieben. Da, basierend auf den Einschränkung bezogen auf Automatisierbarkeit und Rechenbedarf, viele bekannte Verfahren nicht zu automatisierten Diagnose anwendbar sind, wird in Abschnitt 4.4 ein automatisierbares Verfahren zur Charakteristikaerkennung und Unterscheidung der Szenarien gezeigt. Das Verfahren

basiert auf der Berechnung von zehn einfachen und normierten Performance-Indizes und einer anschließenden Mustererkennung [71]. Jedes Szenario weist aufgrund seiner Charakteristika ein typisches Muster der zehn Indizes auf. Dieses wird zum Ende des Kapitels gezeigt.

4.2 Beschreibung der gängigsten Regelungsszenarien

Aggressive Reglereinstellung

Ein falsch eingestellter Regler oder eine falsch gewählte Regelungsstruktur sind, neben externen Störungen, das in der Industrie wohl am häufigsten vertretene Szenario. Die Beurteilung der gewählten Regelungsstruktur, z.B. Kaskadenregelung oder Störgrößenaufschaltung, soll aufgrund der notwendigen Komplexität nicht Teil der vorliegenden Arbeit sein. Für einen falsch eingestellten Regler gilt jedoch, dass die Einstellung entweder zu träge oder zu aggressiv gewählt wurde. Die träge Regelung weist aufgrund der verzögerten Eingriffsmöglichkeit zwar gewisse Sicherheitsrisiken und ungenutzte Ressourcen auf, führt allerdings stets zu einer stationären und stabilen Regelung. Im Gegensatz dazu arbeitet eine aggressive Regelung permanent an den Grenzen ihrer Stabilität, was oft zu dem Charakteristikum führt, dass die Regelgröße sinusförmig zu oszillieren beginnt. Ist die Regelgröße jedoch durch spezielle Anforderungen (z.B. den Flammpunkt des Mediums) durch einen kritischen Wert y_{krit} zu begrenzen, so wird im industriellen Einsatz (z.B. Petrochemie) der Sollwert häufig unterhalb der optimalen Regelgröße y_{opt} gehalten, um aufgrund der Oszillation, ein Überschreiten des kritischen Werts zu vermeiden.

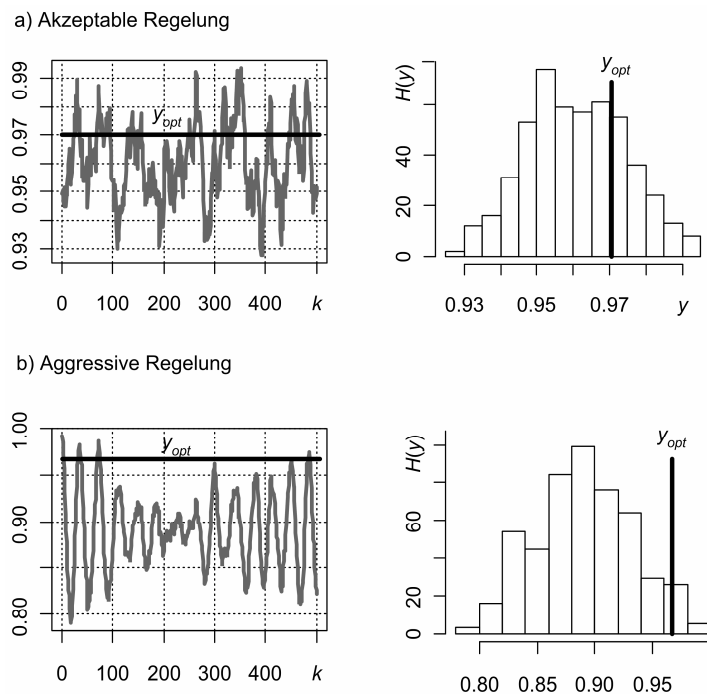


Abbildung 4.1: a) akzeptable und b) aggressive Regelung auf optimalen Wert $y_{opt} = 0.97$ und kritischem Wert $y_{krit} = 1$

In Abbildung 4.1 ist dies für einen Prozess erster Ordnung mit Totzeit $\left(G_p(s) = \frac{1}{0.75s+1} \cdot e^{-s0.25}, \Delta T = 0.05 \right)$ einmal mit einer akzeptablen Regelung ($K_{PR} =$

1.5, $T_N = 1.2$) und einer aggressiven Regelung ($K_{PR} = 2.3$, $T_N = 0.3$) gezeigt. Beide Prozesse werden durch weißes Rauschen mit $\varepsilon \sim N(0, 2.5 \cdot 10^{-5})$ gestört. Als optimale Regelgröße sei $y_{opt} = 0.97$ und als kritischer Wert sei $y_{krit} = 1$ anzunehmen. Unter Annahme, dass der kritische Wert mit einer Wahrscheinlichkeit von 0.995 nicht erreicht wird, lässt sich im Falle der akzeptablen Regelung der Sollwert zu 0.96 wählen. Dahingegen darf bei der aggressiven Regelung und denselben Bedingungen der Sollwert nur zu 0.89 gewählt werden. Neben diesem Gesichtspunkt führt eine aggressive Reglereinstellung auch immer zu einem erhöhten Stellgeräteverschleiß.

In Abbildung 4.2 ist ein skaliertem industrieller Datensatz dargestellt, der das beschriebene Szenario mit dem Charakteristikum der sinusförmigen Oszillation zeigt. Zu sehen sind in a) die Regelgröße, b) der Regelfehler, c) das Stellsignal und in d) die Autokorrelationsfunktion des Regelfehlers.

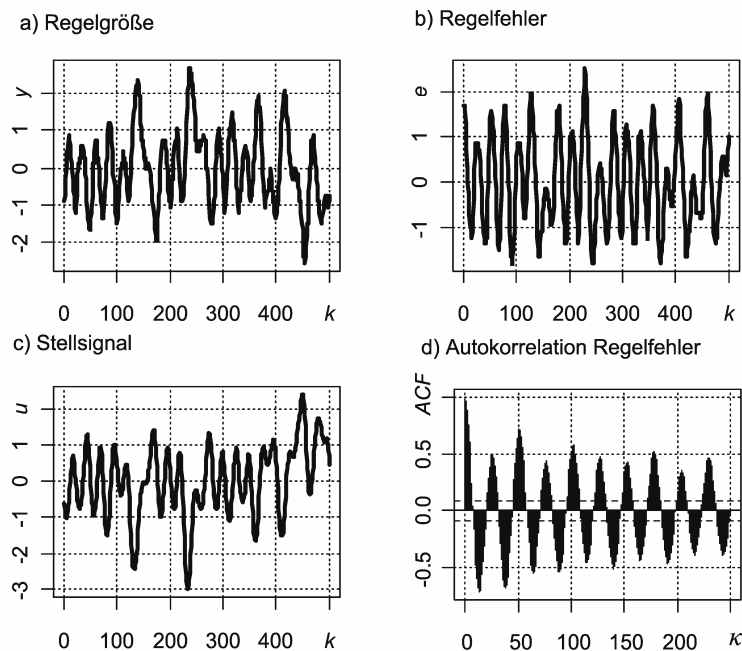


Abbildung 4.2: Aggressive Industrieregung, a) Regelgröße, b) Regelfehler, c) Stellsignal, d) Autokorrelationsfunktion des Regelfehlers

Schwingender Sollwert

Das Szenario „schwingender Sollwert“ basiert in vielen Situationen auf einer aggressiven, überlagerten Regelung, wie es beispielsweise bei der Kaskadenregelung vorkommen kann. Dies führt, genau wie im Szenario der aggressiven Regelung zu dem Charakteristikum der oszillierenden Regelgröße, allerdings mit dem Unterscheidungsmerkmal, dass im Falle der aggressiven Regelung der Sollwert in aller Regel konstant ist oder zumindest nicht oszilliert. In Abbildung 4.3 ist ein Industriebeispiel für dieses Szenario dargestellt. Anhand der Autokorrelationsfunktionen von Sollwert und Regelfehler wird deutlich, dass in diesem Szenario beide Funktionen periodisch oszillieren.

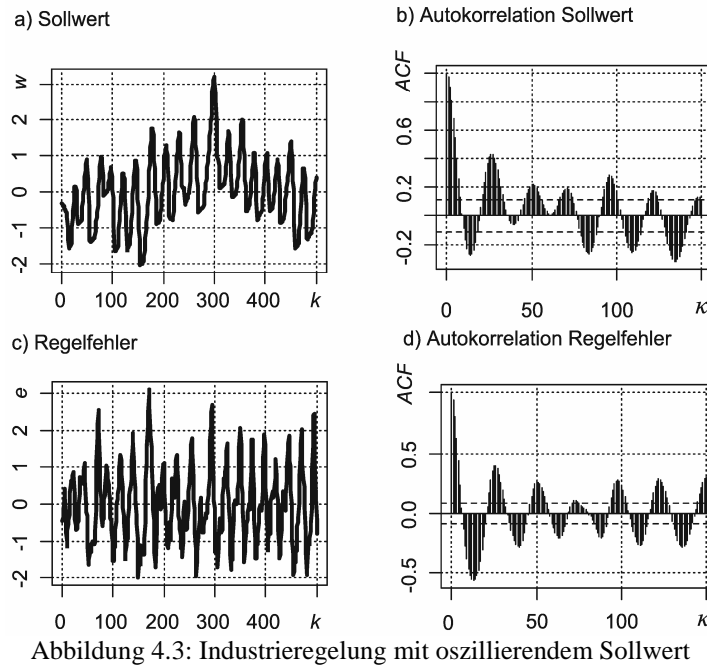


Abbildung 4.3: Industrieregung mit oszillierendem Sollwert

Messrauschen

Die Größe des Messrauschens ε_k ist bei stationärer Regelung für die Stellgeräteaktivität verantwortlich. Sie kann, je nach eingesetztem Reglertyp, stark variieren. So führt der differenzierende Anteil eines Reglers stets zu einer erhöhten Stellgeräteaktivität und somit auch zu einem erhöhten Stellgeräteverschleiß. Das ist in Abbildung 4.4 für einen Prozess erster Ordnung mit Totzeit $\left(G_p(s) = \frac{1}{0.75s+1} \cdot e^{-s0.25}, \Delta T = 0.05 \right)$ und einem PI- ($K_{PR} = 0.75, T_N = 1.2$) und einem PID-Regler ($K_{PR} = 0.75, T_N = 1.8, T_V = 0.6$) gezeigt. Für das Messrauschen gilt $\varepsilon \sim N(0, 1 \cdot 10^{-4})$. Es ist erkennbar, dass bei ähnlichem Regelverhalten der PID-Regler eine stark gestiegene Stellgeräteaktivität aufweist.

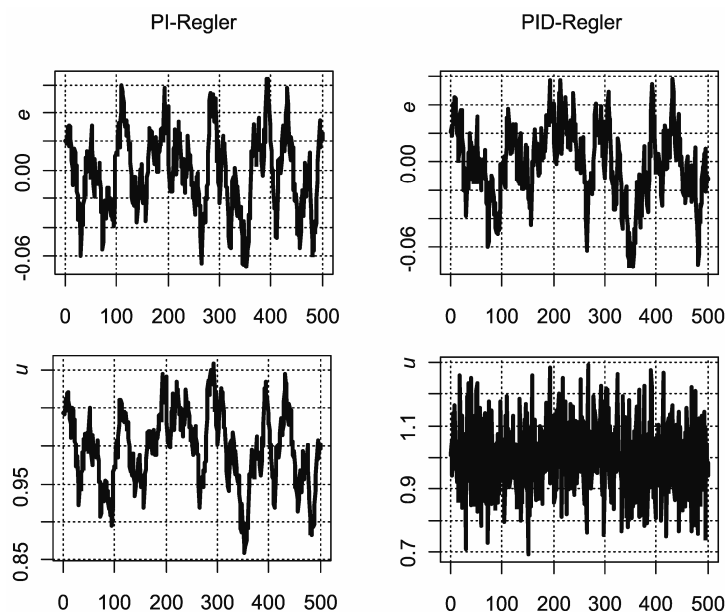


Abbildung 4.4: Einfluss des Reglertyps (PI und PID) auf die Stellgeräteaktivität

Begrenzungen

Von einer begrenzten Regelung wird gesprochen, wenn entweder Stellgeräte oder Prozesse derart falsch ausgelegt werden, dass der Sollwert durch die Regelung nicht erreicht werden kann. Das Charakteristikum dieses Szenarios ist ein stationärer Regelfehler, der sich signifikant von Null unterscheidet. In Abbildung 4.5 ist ein industrieller Datensatz gezeigt, der dieses Szenario während den ersten ca. 50 Abtastpunkten mit dem genannten Charakteristikum aufweist.

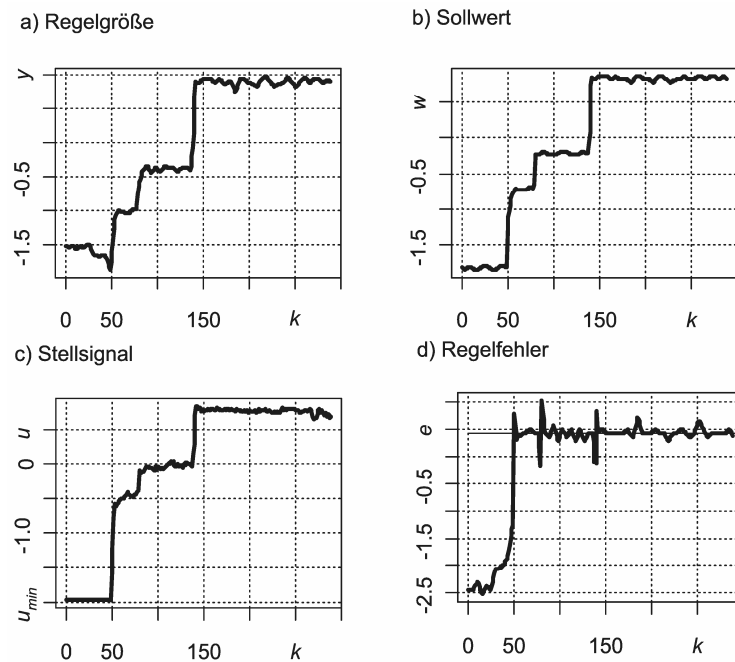


Abbildung 4.5: Begrenzte Industrieregulation während den ersten 50 Abtastpunkten

Neben den Risiken und wirtschaftlichen Schaden, die sich direkt aus dem genannten Charakteristikum der begrenzten Regelung ergeben, kommt es bei diesem Szenario häufig zusätzlich zum so genannten *Windup*. Hierbei steigt der Integrator aufgrund des konstanten Regelfehlers linear mit der Zeit an. Für den Operator sind die Auswirkungen des Windups während dieser Zeit nicht erkennbar, nimmt der Sollwert allerdings wieder Werte an, welche von der Regelung erreicht werden können, so gibt es einen Zeitverzug bis das Stellsignal dem geänderten Sollwert folgt. Im industriellen Einsatz werden hierfür entweder Anti-Windup Filter (s. [21]) eingesetzt oder aber es wird der Integrator angehalten, sobald das Stellsignal seine Begrenzungen erreicht hat.

Inaktive Regelung

Wird ein Regler durch externe Störungen oder einen geänderten Sollwert an seine Stabilitätsgrenze gebracht, so ist das Abschalten des Reglers die in der Industrie am weit verbreitetste Gegenmaßnahme. Der Grund hierfür liegt darin, dass das Eliminieren der externen Störung oder die Neueinstellung des Reglers in vielen Fällen zu kompliziert und zeitintensiv wäre. Realisiert wird das Abschalten durch die manuelle Vorgabe des Stellsignals. Weisen historische Daten eine häufige Abschaltung der Regelung auf, so verbirgt sich in vielen Fällen ein ernsteres Regelungsszenario dahinter. In Abbildung 4.6 ist ein industrieller Datensatz gezeigt. Bis zum 120. Abtastpunkt schwingt die Regelgröße und stabilisiert sich nach Abschalten der Regelung. Nach etwa

300 Abtastpunkten wird die Regelung erneut eingeschaltet, was jedoch wieder zu einer Schwingung der Regelgröße führt.

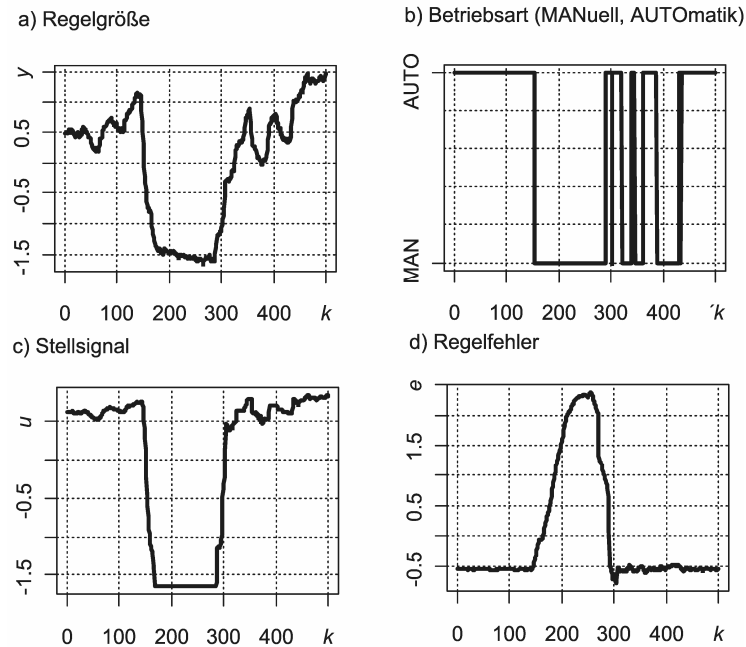


Abbildung 4.6: Inaktive Industrieregung in Folge einer instabilen Regelung

Die Charakteristika der inaktiven Regelung sind ein konstantes Stellsignal und ein stationärer Regelfehler, der sich signifikant von Null unterscheidet.

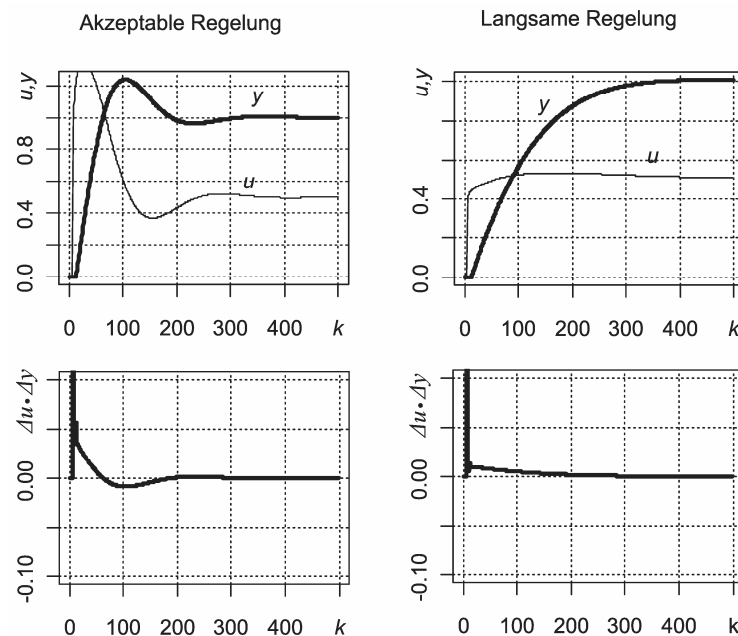


Abbildung 4.7: Gleichgerichtetheit (unten) einer langsamen und akzeptablen Regelung (oben)

Träge Reglereinstellung

Eine träge Reglereinstellung ist weitaus weniger kritisch als eine zu aggressive Reglereinstellung. Sie kommt im industriellen Umfeld jedoch gelegentlich vor. Nach [24] ist das Charakteristikum einer langsamen Regelung die Gleichgerichtetheit zwischen Stellsignal und Regelgröße. In Abbildung 4.7 ist für einen Prozess erster

Ordnung mit Totzeit $\left(G_p(s) = \frac{2}{5s+1} \cdot e^{-s \cdot 0.25}, \Delta T = 0.05 \right)$ eine langsame ($K_{PR} = 0.4$, $T_N = 3.9$) und eine akzeptable Regelung mit einer maximalen Überschwungung von 20% ($K_{PR} = 1$, $T_N = 1.4$) nach einem Sollwertsprung von 0 auf 1 gezeigt. Der simulierte Prozess ist frei von Messrauschen. Um die Gleichgerichtetheit zwischen Stellsignal Regelgröße zu visualisieren, schlägt [24] das Produkt aus $u_k - u_{k-1}$ und $y_k - y_{k-1}$ vor. Ist dieses Produkt größer Null, so bewegen sich Stellsignal und Regelgröße in dieselbe Richtung.

Externe Störungen

Sofern externe Störungen messbar sind, lässt sich deren Einfluss leicht abschätzen, etwa durch die Bestimmung deren Varianz. Außerdem führen oszillierende externe Störungen, ohne geeignete Reglerstruktur, zu einer oszillierenden Regelgröße. Daher kann die externe Störung, sofern sie bekannt ist, ebenfalls auf Oszillation untersucht werden. Im industriellen Umfeld sind viele externe Störungen jedoch nicht messbar. In Abbildung 4.8 ist ein skaliertes industrieller Datensatz gezeigt, der durch eine messbare, externe Störung ε beeinflusst wird. Das Charakteristikum dieses Szenarios besteht darin, dass bis zu einem gewissen Zeitpunkt die stationäre Regelgröße bei gleichzeitig konstantem Sollwert instationär wird. Um zu zeigen, dass in Abbildung 4.8 der Sollwert stationär ist, wurde der Sollwert mit demselben Skalierungsfaktor multipliziert wie die Regelgröße.

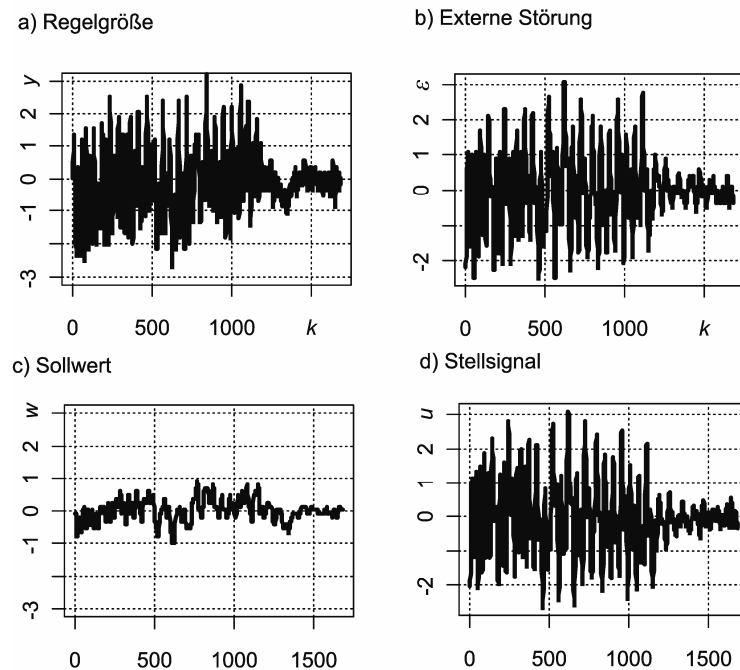


Abbildung 4.8: Industrieregung mit einer externen Störung

4.3 Beschreibung des Regelungsszenarios Haft- und Gleitreibung

Nach [29] bestehen chemische Anlagen aus durchschnittlich 2000 Regelkreisen. Viele dieser Regelkreise nutzen als Stellgerät Regelventile, die sich in ihrem Aufbau und ihrer Wirkungsweise zwar stark unterscheiden können, jedoch wird jedes von ihnen durch Haft- und Gleitreibung im Ventilkörper beeinflusst. Der Grund hierfür liegt darin, dass ein Ventil aus aneinander reibenden Flächen besteht, wodurch dieses Phänomen stets hervorgerufen wird. Kritisch wird es allerdings häufig erst, wenn die Haft- und Gleitreibung durch Ablagerungen (z.B. Rost oder gefördertes Produkt), Schäden an Packungen oder aufeinander reibende Metallflächen verstärkt wird. In Abbildung 4.9 ist die Darstellung eines pneumatischen Ventils aus [38] und [44] dargestellt. Dort kann Haft- und Gleitreibung im Bereich der Packung oder aber auch zwischen Verschluss und Ventilsitz hervorgerufen werden. Durch Ändern des Stellsignals ändert sich bei dem gezeigten Ventiltyp die Ventilöffnung durch den Hub des Verschlusskolbens.

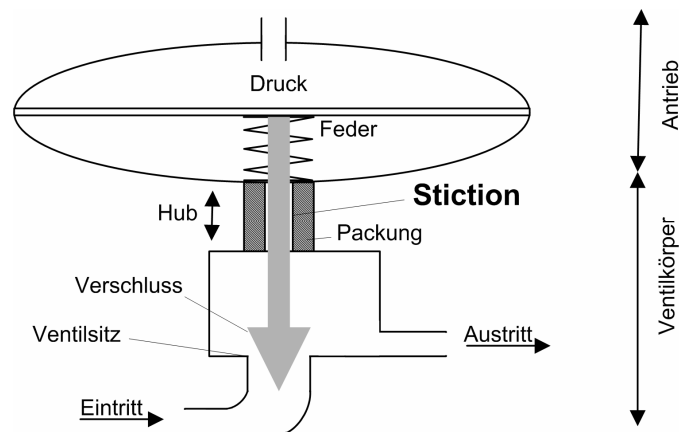


Abbildung 4.9: Darstellung eines pneumatischen Ventils [44]

Das Regelungsszenario Haft- und Gleitreibung zeichnet sich durch das Charakteristikum aus, dass die Regelgröße zu schwingen beginnt. Auswertungen ([9] und [15]) zeigen, dass 20 bis 30% aller industriellen Regelkreise, die ein Regelventil beinhalten, eine schwingende Regelgröße aufweisen. Der Grund für die hervorgerufene Schwingung lässt sich durch die Nichtlinearitäten im Ventilkörper erklären. Die Regelung reagiert durch die abrupten Bewegungen des Verschlusskolbens, wie sie nach dem Überwinden der Haftreibung entstehen, sehr viel aggressiver. Eine ausführlichere Beschreibung dafür ist am Ende dieses Abschnitts zu finden. Dies führt stets zu einer erhöhten Streuung der Regelgröße und zu einem steigenden Verschleiß des Regelventils, was wiederum die Haft- und Gleitreibung fördert. Das korrekte Diagnostizieren eines solchen Regelungsszenarios ist somit von eminenter Wichtigkeit.

In aller Regel kommt es im industriellen Umfeld neben der eigentlichen Haft- und Gleitreibung in Regelventilen auch noch zu weiteren, aber sehr verwandten Effekten, die [38] entnommen sind:

- **Hysterese** – Wird das Ventil auf- und zugefahren, so kommt es im Falle einer Hysterese zu zwei verschiedenen Verläufen. Das bedeutet, dass die Ventilöffnung beim Stellsignal u_k während des Auffahrens eine andere Ventilöffnung $u_{v,k}$ besitzt als beim Zufahren.
- **Totband** – Bei einer Änderung der Fahrtrichtung (auf- oder zufahren) kommt es häufig zum Totband. Eine Änderung des Stellsignals führt in diesem Bereich nicht zu einer Änderung der Ventilöffnung. Der Grund hierfür liegt ebenfalls in einem gewissen Spielraum zwischen den sich bewegenden Teilen. In Abbildung 4.10 ist der Grundgedanke gezeigt für zwei Zahnräder. Beim Ändern der Bewegungsrichtung des angetriebenen Zahnrads 1 muss zunächst das Totband überwunden werden, bevor Zahnrad 2 sich bewegt. Das Totband verursacht eine Phasenverschiebung zwischen dem Eingangssignal (Stellsignal) und dem Ausgangssignal (Ventilöffnung).
- **Totzone** – Im Vergleich zum Totband ist die Totzone nicht richtungsabhängig, sondern abhängig von der Größe des Stellsignals. Es kommt daher nicht zu einer Phasenverschiebung.

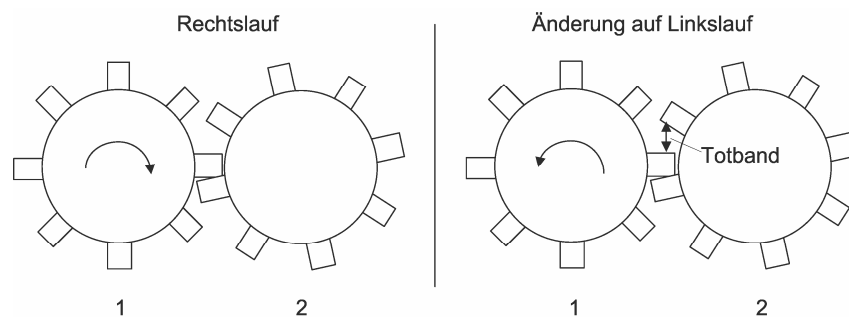


Abbildung 4.10: Darstellung des Totbands bei Zahnradantrieb

Alle diese Begriffe werden im englischsprachigen Raum allgemein als *Stiction* (*Static friction*) zusammengefasst. Um das Szenario eindeutig durch die Fehlerdiagnose erkennen zu können, muss somit ein Verfahren zur Simulation von Stiction herangezogen werden. Dafür gibt es zahlreiche Verfahren, die in [38] zusammengetragen sind. Die Verfahren können entweder physikalischer Art sein, wie [17], [40] und [65], oder datenbasiert, wie [5] oder [39]. Physikalische Modelle sollen aufgrund der Komplexität, hervorgerufen durch die Vielzahl an Einstellparameter wie Federkonstanten oder Reibkoeffizienten, nicht verwendet werden. Da sich nach [38] die beiden datenbasierten Verfahren kaum unterscheiden, wird das Verfahren nach [5] im Folgenden zur Simulation verwendet. Das Verfahren nutzt die beiden Parameter S und J zur Angabe der Ausprägung von Stiction. S entspricht dabei dem Stellsignalsbereich (in %), der nach Richtungsänderung überwunden werden muss, bevor die Ventilöffnung sich ändert. Der Parameter J gibt die Sprungweite an, welche die Ventilöffnung nach Überwinden der Haftreibung vollzieht. Das Eingang-Ausgangsverhalten eines solchen Ventilmodells ist in Abbildung 4.11 dargestellt.

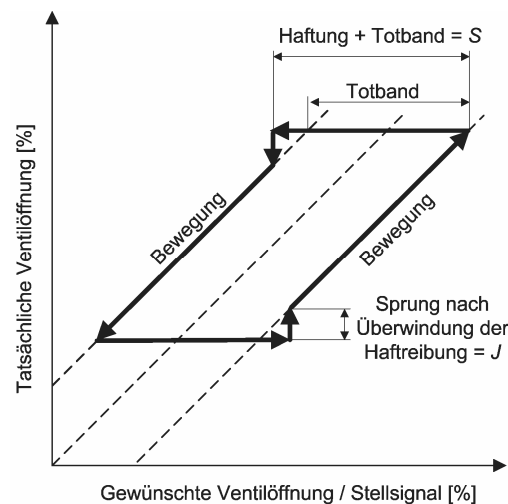


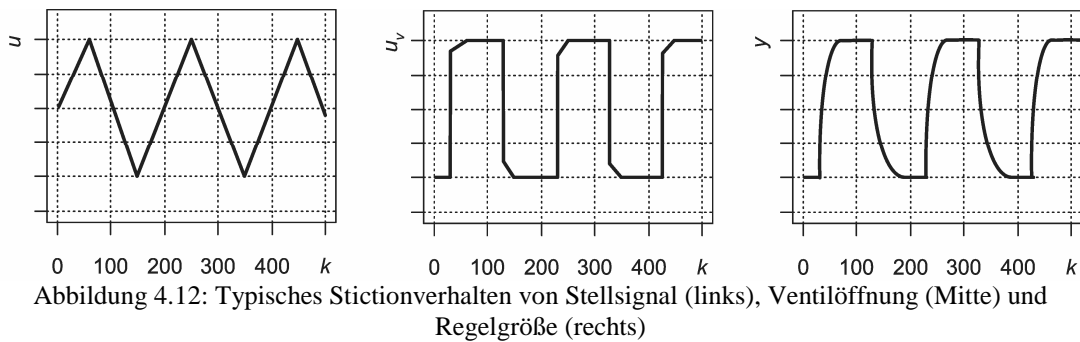
Abbildung 4.11: Ein- und Ausgangsverhalten bei Stiction Modell nach [5]

Das Erkennen des Regelungsszenarios „Haft- und Gleitreibung“ ist circa seit dem Jahre 1995 Aufgabe vieler Forschungsgruppen und Gegenstand vieler internationaler Veröffentlichungen. Nach [38] lassen sich die wichtigsten davon jedoch in die folgenden Gruppen einteilen:

- Spezielle Verfahren der Oszillationserkennung
- Analyse der Phasenverschiebung zwischen Stellsignal und Regelgröße
- Nichtlineare Optimierungen
- Verfahren basierend auf Mustererkennung

Die meisten Verfahren sind dabei der Gruppe „Mustererkennung“ zuzuordnen, so z.B. [30] (patentiert durch [31]) der die Verteilung der Regelgrößenänderung analysiert, [26] der sowohl Stellsignal als auch Regelgröße durch verschiedene Kurvenzüge (Sinus, Dreieck, Rechteck) darstellt und die Ähnlichkeit damit überprüft, [51] der den tatsächlichen Regler durch einen optimierten Zweipunktregler ersetzt und die Ähnlichkeit zwischen simuliertem Regelgrößenverlauf und tatsächlichem Regelgrößenverlauf analysiert. Ein weiteres Verfahren aus dieser Gruppe wird im Folgenden eingeführt [75].

Zahlreichen Quellen ist zu entnehmen (z.B. [18] und [55]), dass Ventile mit Haft- und Gleitreibung, welche durch einen PI-Regler in Prozessen mit Ausgleich geregelt werden, Oszillationen mit einer ganz speziellen Form hervorrufen. Diese Form wird im Allgemeinen als Rechtecksverlauf bezeichnet. Die Oszillation ist dadurch zu begründen, dass die Ventilöffnung beim Ansteigen des Stellsignals ab einer gewissen Grenze das Losbrechmoment überwindet und schlagartig auffährt. Dasselbe geschieht, wie auch Abbildung 4.11 zu entnehmen ist, auch beim Schließen des Ventils. Die Ventilstellung besitzt somit im Grunde lediglich die binären Zustände „zu weit offen“ oder „zu weit geschlossen“ zwischen denen sie periodisch pendelt. Die typische Rechtecksform kommt somit durch die Sprungantworten des Systems und durch die Phasen der unüberwundenen Haftreibung zustande. In Abbildung 4.12 sind die Verläufe des Stellsignals, der Ventilöffnung u_v und der Regelgröße qualitativ dargestellt.



Im Vergleich dazu resultieren eine aggressive Reglereinstellung oder externe Störungen häufig in einer eher sinusförmigen Schwingung. Diese Tatsache kann genutzt werden indem diese beiden Formen der Regelgröße (sinusförmig und quasi-rechteckförmig) mittels PCA (*Principal Component Analysis*, Hauptkomponentenanalyse) unterschieden werden. Das genannte Verfahren besteht, wie jede andere Mustererkennung ebenfalls, aus den Schritten der Rohdatenvorbehandlung, des Trainings und der eigentlichen Klassifikation.

Der Erfolg der Klassifikation, also der Entscheidung, ob es sich um einen sinusförmigen oder einen quasi-rechteckförmigen Verlauf handelt, hängt vom Umfang und der Qualität der Trainingsdaten ab. Trainingsdaten sind dabei Daten, deren Zugehörigkeit eindeutig bekannt ist. Die Generierung von Trainingsdaten kann dabei entweder durch Simulation (z.B. durch die Verwendung des Stiction-Modellansatzes nach [5]) erfolgen oder aber durch die erfolgreiche Zuordnung historischer Betriebsdaten anhand der genannten Verfahren zur Erkennung von Haft- und Gleitreibung. Das bedeutet, dass historische Daten aus dem zu untersuchenden Prozess manuell auf Stiction oder Sinusschwingung untersucht werden. Der Nachteil dieses Ansatzes ist der enorme Zeitaufwand. Im Gegensatz dazu besteht der Nachteil bei der Verwendung von Simulationsdaten darin, dass die Mustererkennung am besten arbeitet, wenn sowohl ein Prozessmodell bekannt ist als auch die Modellparameter S und J plausibel gewählt wurden. Zunächst scheint dieses Verfahren nicht dem Wunsch der Automatisierbarkeit zu entsprechen, jedoch handelt es sich bei der Generierung von Trainingsdaten (für jeden einzelnen Prozess) um eine einmalige Tätigkeit. Die restlichen Schritte der Erkennung sind völlig automatisierbar. In [75] wurde gezeigt, dass das Verfahren auch bei der Verwendung von simulierten Trainingsdaten und nur einer groben Annahme des Prozessmodells zuverlässig funktioniert.

Um dies für einen nahezu unbekannten Prozess zu zeigen, wurden dabei 50 Simulationen eines Trainingsprozess erster Ordnung ohne Totzeit mit dem genannten Ventilmodell durchgeführt. Für die Modellparameter galt dabei, dass

$$S, J \sim N(20, 25), \quad (4.1)$$

um eine möglichst große Spanne an Ausprägung erkennen zu können. Um Vergleichsdaten des Musters „sinusförmig“ zu erhalten, wurden als weitere Trainingsdaten Sinusschwingungen mit unterschiedlicher Rauschintensitäten simuliert.

Da sich Trainingsdaten auch noch im Hinblick auf Abtastzeit, Zeitkonstanten und Amplitude der Schwingung von den Untersuchungsdaten unterscheiden können, müssen sowohl Trainingsdaten als auch Untersuchungsdaten vorbehandelt werden. Diese Vorbehandlung umfasst die folgenden Punkte:

- Ausschließliche Verwendung von Messabschnitten, in denen der Sollwert über einen langen Zeitraum konstant ist.
- Skalierung auf den Mittelwert 0 und die Varianz 1.
- Separierung der einzelnen Perioden (zwei Nulldurchgänge), die jedoch eine minimale Länge aufweisen müssen, um Rauschen auszuschließen.
- Interpolation auf eine feste Anzahl an Messwerten pro Periode (im Folgenden 100 Werte).

Zum besseren Verständnis ist die Vorgehensweise vom ursprünglichen Datensatz hin zum vorbehandelten Muster in Abbildung 4.13 gezeigt.

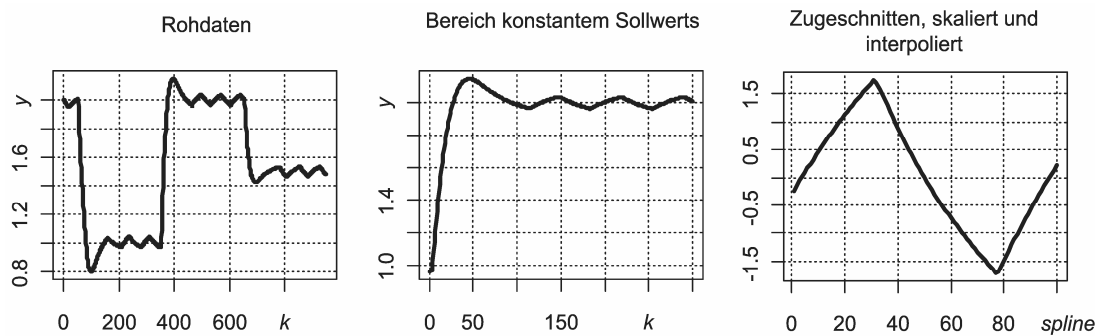


Abbildung 4.13: Schritte der Rohdatenvorbehandlung

Sind alle Trainingsdaten auf diese Weise vorbehandelt, so werden sie im nächsten Schritt zeilenweise zu einer Matrix zusammengefügt. Unter der Annahme, dass bei der Interpolation n_s Werte erzeugt wurden, so entspricht nun, bei zeilenweiser Betrachtung, jedes Muster einem Punkt im n_s dimensionalen Raum. Die Erzeugung der Trainingsdaten ist damit abgeschlossen.

Im letzten Schritt, der Klassifikation, wird das vorbehandelte Untersuchungsmuster (im Betrieb gemessen) ebenfalls als ein Punkt im n_s Dimensionalen betrachtet. Die Zuordnung „sinusförmig“ oder „quasi-rechteckförmig“ wird anhand der Entfernungen zu den beiden Punktwolken getroffen. Dies kann beispielsweise mithilfe der k-nächste-Nachbar-Klassifikation [7] erfolgen. Nachteil dieser Herangehensweise ist, dass das Ergebnis nicht grafisch darstellbar ist (somit wenig benutzerfreundlich) und stark durch Rauschen und Ausreißer beeinflusst wird. Beide Nachteile können umgangen werden, wenn der n_s -dimensionale Raum mittels PCA auf einen visualisierbaren zweidimensionalen Unterraum herunter projiziert wird. Dadurch wird weiterhin die wichtigste und größte Information dargestellt und ist zusätzlich bedienerfreundlicher als der höher dimensionale Ansatz. In Abbildung 4.14 sind für simulierte Trainingsdaten die ersten beiden Hauptkomponenten (PC_1 und PC_2) gegeneinander aufgetragen, die zusammen nahezu 99% der gesamten Information (Varianz) wiedergeben. Dargestellt sind ebenfalls die berechneten Konfidenzbereiche, in denen 95% aller Messdaten liegen. Die beiden Gruppierungen sind in diesem Raum klar trennbar.

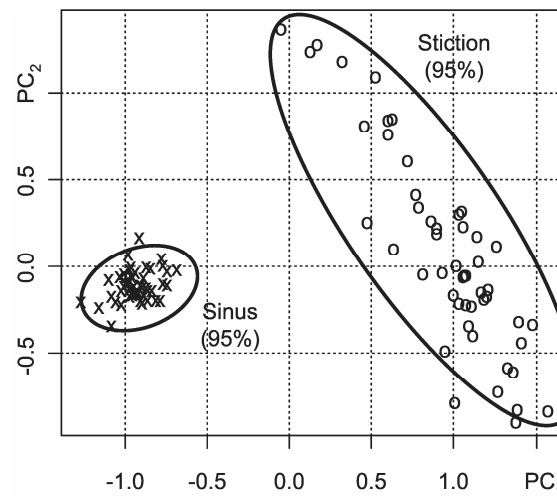


Abbildung 4.14: Darstellung der Trainingsdaten im Raum der ersten beiden Hauptkomponenten [75]

Bei der anschließenden Klassifizierung wird anhand der berechneten *Loadings* (Abbildungsmatrix, die bei der Diagonalisierung der Kovarianzmatrix anfällt) das Muster des zu untersuchenden Datensatzes ebenfalls in den zwei dimensional Unterraum abgebildet. Für das Muster aus Abbildung 4.15 (links), mit anzunehmender unbekannter Zugehörigkeit (simuliert wurde Stiction), ergibt sich die Abbildung im PC_1 - PC_2 Raum aus Abbildung 4.15 (rechts). Es ist erkennbar, dass dieser Datensatz somit das typische Muster einer Stictionschwingung besitzt.

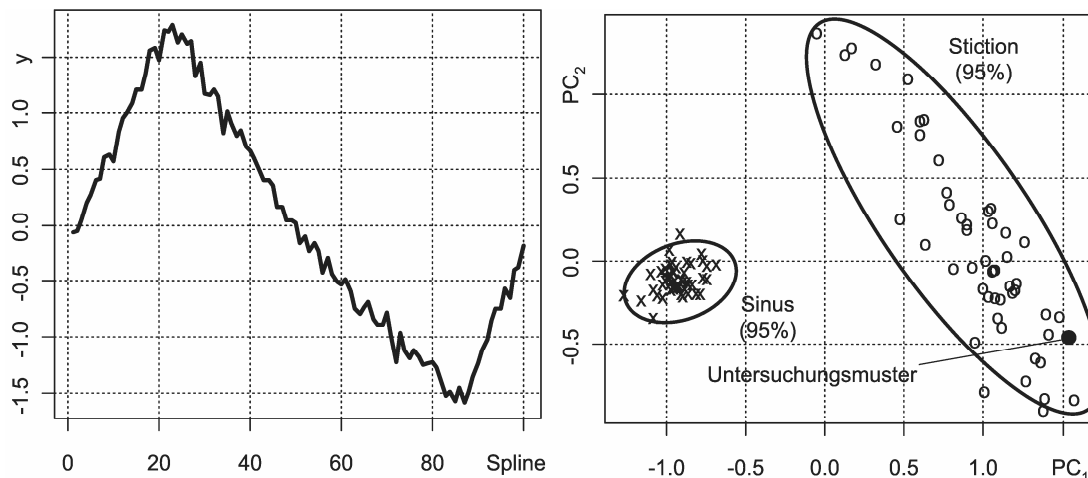


Abbildung 4.15: Simuliertes Muster mit unbekannter Zugehörigkeit (links) und dessen Abbildung in den Unterraum der Trainingsdaten [75]

Um zu zeigen, dass die Mustererkennung, basierend auf simulierten Trainingsdaten, ebenfalls ohne genaue Kenntnis des zugrunde liegenden Prozesses zu einer hohen Erkennungsrate führt, wurde sie zum einen für Prozesse getestet, welche eine andere Ordnung besitzen, als jener der zur Generierung der Trainingsdaten verwendet wurde. Der Trainingsprozess ist erster Ordnung ohne Totzeit und wird für die Prozesse „erster Ordnung ohne Totzeit“, „erster Ordnung mit Totzeit“, „dritter Ordnung ohne Totzeit“ und „dritter Ordnung mit Totzeit“ getestet. Die Übertragungsfunktionen, Modellparameter und Erkennungsrate werden in Tabelle 4.1 dargestellt.

Tabelle 4.1: Erkennungsrate bei verschiedenen Prozessen [75]

Simulationsparameter	Erkennungsrate
$G_p(s) = \frac{2}{1+10s}$, $S = J = 20$, $\sigma_\varepsilon = 0.2\%$	96%
$G_p(s) = \frac{2}{(1+1/3s)^3}$, $S = J = 20$, $\sigma_\varepsilon = 0.2\%$	86%
$G_p(s) = \frac{2}{1+10s} e^{-2s}$, $S = J = 20$, $\sigma_\varepsilon = 0.2\%$	92%
$G_p(s) = \frac{2}{(1+1/3s)^3} e^{-2s}$, $S = J = 20$, $\sigma_\varepsilon = 0.2\%$	84%

Außerdem soll das Verfahren für einen industriellen Datensatz getestet werden. Das vorbehandelte Untersuchungsmuster ist in Abbildung 4.16 dargestellt auch mit seiner Lokalisierung im PCA-Raum. Das Muster liegt zwar nah an der Klassifikationsgrenze, jedoch kann es, unter Berücksichtigung des Konfidenzbereichs, noch der Gruppe *stiction* zugeordnet werden.

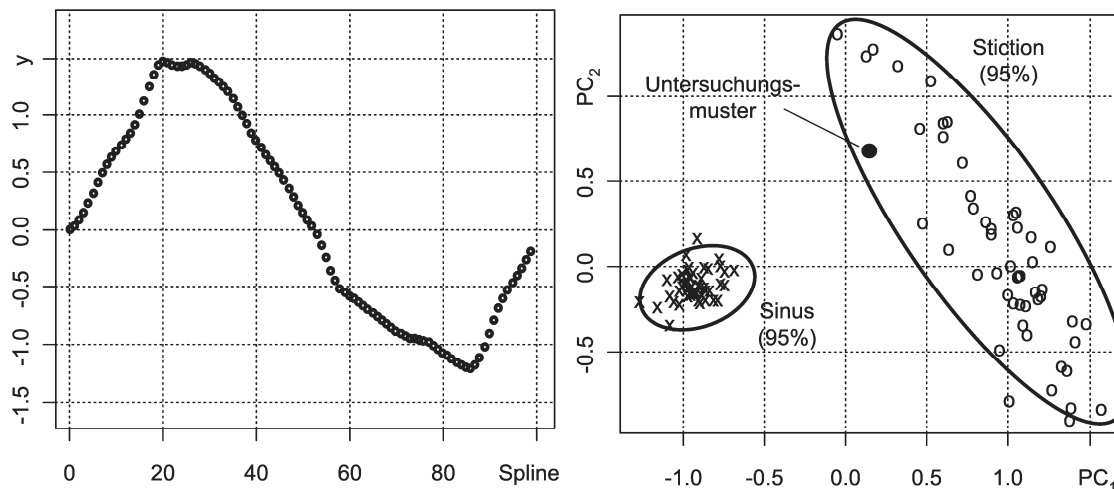


Abbildung 4.16: Industrielles Muster mit unbekannter Zugehörigkeit (links) dessen Abbildung in den Unterraum der Trainingsdaten [75]

Es wurde in [75] mit Hilfe anderer Verfahren ([27], [28]) zur Diagnose von Haft- und Gleitreibung gezeigt, dass diese Entscheidung vermutlich korrekt ist.

Zusammengefasst lässt sich sagen, dass die Güte der Identifizierung von der Qualität der Trainingsdaten abhängt. Werden diese Trainingsdaten durch Simulationen generiert, so werden optimale Ergebnisse erzielt, wenn ein genügend gutes Prozessmodell vorliegt. Es wird vermutet, dass sofern Trainingsdaten aus historischen, erfolgreich klassifizierten Daten gewonnen werden, das Resultat bei dieser Art der Erkennung am besten ist.

4.4 Automatisierte Fehlerdiagnose einzelner Regelkreise

Basierend auf den in den Abschnitten 4.2 und 4.3 vorgestellten Regelungsszenarien mit den daraus abgeleiteten Charakteristika lässt sich die Zusammenfassung aus Tabelle 4.2 ableiten.

Tabelle 4.2: Regelungsszenarien mit den entsprechenden Charakteristika

Szenario	Charakteristikum
Aggressive Reglereinstellung	Regelgröße schwingt sinusförmig bei konstantem Sollwert und nicht oszillierenden externen Störungen (sofern messbar)
Schwingender Sollwert	Sollwert schwingt
Messrauschen	Regelgröße weist stochastischen Anteil auf
Begrenzungen	Stellsignal ist konstant und arbeitet an seinen Grenzen. Außerdem verhält sich der Regelfehler stationär bei gleichzeitiger signifikanter Abweichung von Null.
Inaktive Regelung	Stellsignal ist über einen langen Zeitraum konstant. Außerdem verhält sich der Regelfehler stationär bei gleichzeitiger signifikanter Abweichung von Null.
Träge Reglereinstellung	Häufige Gleichgerichtetheit zwischen Regelgröße und Stellsignal nach Sollwertänderung.
Haft- und Gleitreibung	Regelgröße und Stellsignal weisen die typische Schwingungsform bei konstantem Sollwert auf.
Externe Störungen	Die (bis zum Zeitpunkt des Auftretens der externen Störung) stationäre Regelgröße wird trotz stationärem Sollwert instationär.

Verschiedene Maßzahlen sollen im Folgenden vorgestellt werden, die es ermöglichen die genannten Charakteristika objektiv zu messen, um somit Aussagen über das zugrunde liegende Szenario zu ermöglichen. Die Maßzahlen werden mit p_1, p_2, \dots, p_{10} bezeichnet. Dabei sollen alle Maßzahlen normiert (zwischen 0 und 1) und möglichst unabhängig vom Prozess (z.B. Arbeitspunkt, Verstärkung) sein, um eine automatisierte Auswertung zu ermöglichen.

Relative Zeit konstanten Stellsignals, p_1

Sowohl die inaktive Regelung, als auch eine begrenzte Regelung zeichnen sich durch einen großen zeitlichen Anteil aus, in dem das Stellsignal konstant ist. Daher soll für eine Datenreihe aus n Messungen die binäre Größe

$$u_{const,i} = \begin{cases} 1, & \text{wenn } u_{i+1} - u_i = 0 \\ 0, & \text{sonst} \end{cases} \quad (4.2)$$

nach [71] berechnet werden. Daraus berechnet sich die relative Zeit stationären Stellsignals p_1 gemäß

$$p_1 = \frac{1}{n-1} \sum_{i=1}^{n-1} u_{const,i} \quad (4.3)$$

Zur Festlegung eines Grenzwerts dienen industrielle Daten aus der Petrochemie eines Reglers, der während seines Betriebs stets eine eingeschaltete Regelung besaß. Die Daten wurden in einem Zeitraum von 45 Tagen (genauer 64803 Minuten) im normalen Betrieb gesammelt. Während 4713 Abtastpunkten gab es keine Stellsignaländerung. Der daraus berechnete p_1 Wert beträgt 0.07. Damit wird der kritische Wert, unter Berücksichtigung einer Sicherheit, zu 0.15 gewählt. Es ist zu beachten, dass diese Berechnung nur anzuwenden ist, wenn das Stellsignal nicht zu stark quantisiert und gefiltert wird.

Relative Zeit stationären und signifikanten Regelfehlers, p_2

Ein Signal gilt auf einem Zeitintervall als stationär, wenn dessen Erwartungswert darauf konstant ist. Die Identifikation stationärer Bereiche eines Signals ist nach [4] durch die Betrachtung der Signalvarianz und der Signaländerungsvarianz möglich. Als Ergebnis lässt sich ein binärer Vektor $s_{sid,x}$ der Länge n für ein Signal x berechnen. Hat dieser an der Stelle k den Wert 1, so ist das untersuchte Signal dort stationär, andernfalls (Wert 0) instationär. Darauf basierend kann die Variable $e_{sist,i}$ nach [71] berechnet werden gemäß

$$e_{sist,i} = \begin{cases} 1, & \text{wenn } s_{sid,e,i} = 1 \wedge \frac{|e_i|}{|w_i|} > 0.1 \\ 0, & \text{sonst} \end{cases} \quad (4.4)$$

Die Bedingung $\frac{|e_i|}{|w_i|} > 0.1$ überprüft, ob der Regelfehler zum Abtastpunkt i mindestens 10% vom aktuellen Sollwert beträgt und somit heuristisch als signifikant betrachtet wird. Daraus lässt sich direkt die relative Zeit stationären und signifikanten Regelfehlers p_2 berechnen gemäß

$$p_2 = \frac{1}{n} \sum_{i=1}^n e_{sist,i} \quad (4.5)$$

Zur Festlegung eines Grenzwerts dienen erneut industrielle Daten aus der Petrochemie eines Reglers, der während seines Betriebs stets eine eingeschaltete und nicht begrenzte Regelung besaß. Die Daten wurden in einem Zeitraum von 45 Tagen (genauer 64803 Minuten) im normalen Betrieb gesammelt. Während 973 Abtastpunkten gab es einen stationären und signifikanten Regelfehler. Der daraus berechnete p_2 Wert beträgt 0.015. Damit wird der kritische Wert, unter Berücksichtigung, einer Sicherheit zu 0.03 gewählt.

Sättigungsindex, p_3

Sowohl begrenzte als auch inaktive Regelungen weisen hohe p_2 und p_3 -Werte auf. Um diese jedoch unterscheiden zu können, wird der Sättigungsindex nach [36] verwendet. Dieser basiert auf der binären Variable $s_{i,i}$, welche angibt, ob das Stellsignal zum Abtastpunkt i an seinen Grenzen arbeitet oder nicht. Mathematisch ist dieser bei einem maximalen (U_{max}) und einem minimalen Stellsignal (U_{min}) definiert als

$$s_{i,i} = \begin{cases} 1, & \text{wenn } \frac{u_i - U_{\min}}{U_{\max} - U_{\min}} < 0.1 \vee \frac{u_i - U_{\min}}{U_{\max} - U_{\min}} > 0.9 \\ 0, & \text{sonst} \end{cases} . \quad (4.6)$$

Daraus lässt sich der Sättigungsindex einer Messreihe bestimmen als

$$p_3 = \frac{1}{n} \sum_{i=1}^n s_{i,i} . \quad (4.7)$$

Im Gegensatz zur inaktiven Regelung, zeichnet sich die begrenzte Regelung durch hohe p_3 -Werte aus. Bei der automatischen Fehlerdiagnose sind die maximalen und minimalen Stellsignale nicht immer bekannt. In diesem Fall sollte vereinfachend angenommen werden, dass $u_{\max} = \max_{i \in [1,n]}(u_i)$ und $u_{\min} = \min_{i \in [1,n]}(u_i)$.

Der Grenzwert wird unter der Annahme der Gleichverteilung des Stellsignals getroffen. Im diesem Idealfall beträgt der theoretische p_3 Wert 0.2. Als kritischer Wert wird, unter Berücksichtigung einer Sicherheit, somit 0.3 vorgeschlagen.

Relative Rauschintensität, p_4

Die Rauschintensität soll der Standardabweichung des Messrauschens σ_e entsprechen. Diese kann, wie in Abschnitt 3.1.1 gezeigt, mithilfe eines geeigneten ARMA Modells geschätzt werden ($\hat{\sigma}_e$). Daraus berechnet sich die relative Rauschintensität p_4 gemäß

$$p_4 = \frac{1}{n} \sum_{i=1}^n \frac{6\hat{\sigma}_e}{|w_i|} . \quad (4.8)$$

Bei der Aufstellung und Auswertung dieses Index wird davon ausgegangen, dass es sich bei der Messstörung ε um weißes Rauschen handelt dessen Streuung während der Datenaufzeichnung konstant ist, so dass $6\hat{\sigma}_e$ die Spannweite von ca. 99.5% aller Messstörungen angibt. Die relative Rauschintensität gibt somit an, wie groß der Rauschanteil bezogen auf den aktuellen Sollwert im Mittel ist. In Abbildung 4.17 ist dieser Index für einen Prozess erster Ordnung mit Totzeit und einem PI-Regler gezeigt. Simuliert wurde dabei zum einen eine Störung mit der Standardabweichung 0.01 und 0.1. Beide Simulationen wurden sowohl für den Sollwert 1 als auch für den Sollwert 1.5 durchgeführt. Die geschätzten Standardabweichungen mit den daraus berechneten p_4 -Werten sind ebenfalls dargestellt.

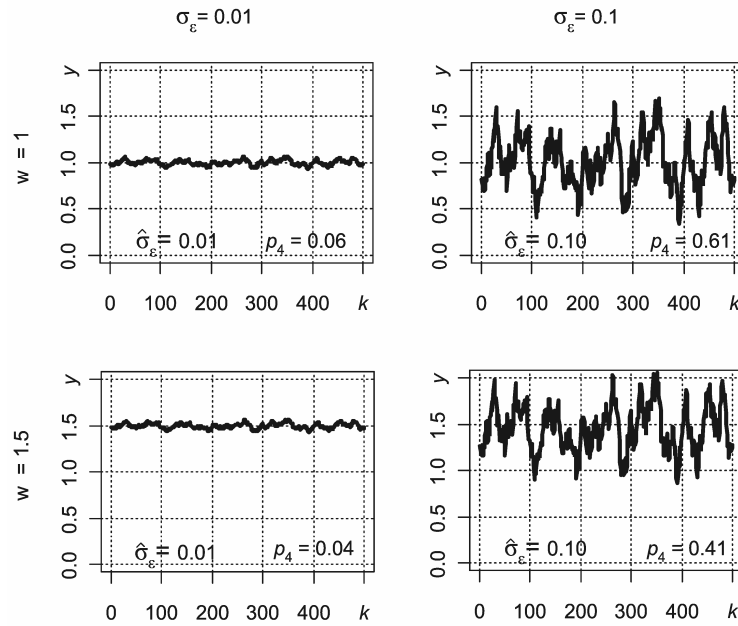


Abbildung 4.17: Verschiedene Rauschintensitäten bei verschiedenen Sollwerten

Bemerkung: Beim Berechnen von relativen Größen, die sich auf andere Prozessgrößen, wie z.B. Sollwert oder Regelgröße beziehen, ist es wichtig diese Bezugsgrößen zuvor eventuell anzupassen um plausible Werte zu erhalten. Dies soll anhand eines einfachen Beispiels erläutert werden: Bei der Temperaturregelung eines offenen Behälter auf den Sollwert 35°C bei einer Umgebungstemperatur von 20°C, ist der Sollwert für die obige Berechnung dementsprechend mit 15°C anzunehmen.

Die Notwendigkeit einer direkten Implementierung der Regelkreisdiagnostik im Prozessleitsystem führt in vielen Fällen dazu, dass die Berechnung eines solchen ARMA-Modells nicht möglich ist. Dann muss ein vereinfachtes, aber auch weniger zuverlässiges Verfahren herangezogen werden [71]. Bei diesem wird die Tatsache genutzt, dass durch mehrfache Ausführung einer diskreten Differenzierung ($y_k - y_{k-1}$) der langsame, durch einen trägen Prozess hervorgerufene Anteil verschwindet oder zumindest nur gering verstärkt wird. Die rapide, durch die stochastische Störung hervorgerufene Signaländerung wird durch das mehrmalige Differenzieren stark verstärkt. So gilt für ein stochastisches Signal x_k mit der Varianz σ_x^2 , dass nach n -maligem Differenzieren ($x^{(n)}$) die Varianz $\sigma_{x^{(n)}}^2$ um den Faktor 2^n gestiegen ist. Der Grund hierfür liegt in der anzunehmenden Unabhängigkeit zwischen x_k und x_{k-1} besteht und somit für die Werte gilt $x_k, x_{k-1} \sim N(\mu_x, \sigma_x^2)$. Dann gilt ferner, dass

$$x_k \pm x_{k-1} \sim N(\mu_x, 2\sigma_x^2) \quad (4.9)$$

und somit auch

$$\sigma_{x^{(n)}}^2 = 2^n \sigma_x^2. \quad (4.10)$$

Wird eine Regelgröße y , welche aus einem stochastischen und einem langsam dynamischen Anteil besteht, somit n -fach diskret differenziert und von diesem Signal die Varianz bestimmt, so lässt sich die Varianz des eigentlichen weißen Rauschens schätzen gemäß

$$\hat{\sigma}_\varepsilon = \sqrt{\frac{\sigma_{y^{(n)}}^2}{2^n}}. \quad (4.11)$$

Der Grund ist darin zu finden, dass die Varianz des dynamischen Anteils weit langsamer zunimmt (oder sogar abnimmt) als die Varianz des stochastischen Anteils. Abweichungen von dieser Überlegung sind dadurch zu begründen, dass zum einen durch das Abtasten des Signals Unstetigkeiten auftreten und zum anderen können Simulationen diese These nur begrenzt belegen, da diese kein ideal weißes Rauschen ($\text{Cov}(x_k, x_{k-1}) \neq 0$) erzeugen können.

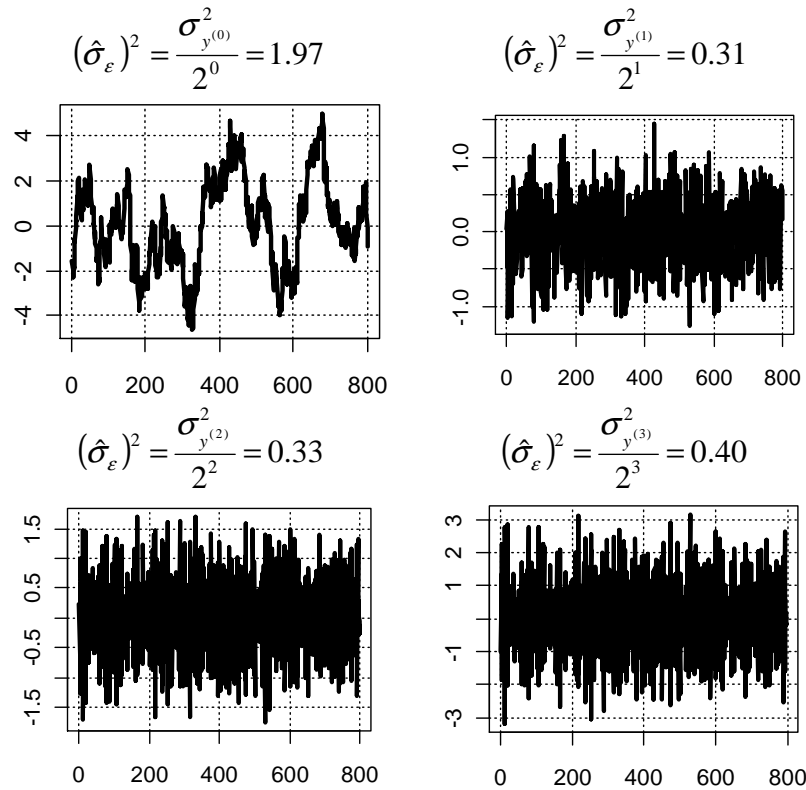


Abbildung 4.18: Dreimalige Differenzierung des Regelfehlers eines geregelten Prozesses, der durch $\varepsilon \sim N(0, 0.45)$ gestört wird mit geschätzter Störungsvarianz $(\hat{\sigma}_\varepsilon)^2$ (oben links: Regelfehler, oben rechts: einmal differenzierter -, unten links: zweimalig differenzierter -, unten rechts: dreimalig differenzierter Regelfehler)

Um die Grundidee jedoch zu zeigen, ist in Abbildung 4.18 die Störgrößenregelung ($K_{PR} = 1.5$, $T_N = 1.2$) eines Prozesses erster Ordnung mit Totzeit $\left(G_p(s) = \frac{1}{3s+1} \cdot e^{-s \cdot 0.15}, \Delta T = 0.05 \right)$ als Reaktion auf Messstörungen mit $\varepsilon \sim N(0, 0.45)$ gezeigt (oben links). In den anderen Darstellungen sind die ersten drei diskreten

Differenzierungen zu sehen, wobei die berechnete Störvarianz jeweils angegeben sind. Es ist zu sehen, dass die geschätzte Varianz gegen die tatsächliche Störvarianz (0.45) strebt. Aufgrund der zeitlichen Korrelation der Messstörungen ($\text{Cov}(\varepsilon_k, \varepsilon_{k-1}) \neq 0$) konvergiert sie jedoch nicht für weitere Differenzierungen sondern steigt langsam weiter an.

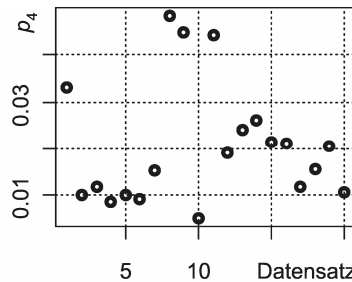


Abbildung 4.19: Berechnete p_4 -Werte für 20 Industriedatensätze

Zur Festlegung eines Grenzwerts soll der verlässlichere ARMA-Ansatz dienen und es werden 20 industrielle Datensätze, die jeweils aus 5000 Messungen mit einminütiger Abtastung bestehen, verwendet. Die Filterordnungen betragen $p = q = 3$. Die berechneten p_4 -Werte sind in Abbildung 4.19 dargestellt. Legt man zugrunde, dass lediglich 5% der Prozesse eine zu hohe Rauschintensität aufweisen sollen, so lässt sich der Grenzwert als ca. 0.05 festlegen.

Gleichverteilung der Stellgröße, p_5

In Abschnitt 4.3 wurde das Phänomen der Haft- und Gleitreibung in Ventilkörpern erklärt und es wurden verschiedene, auch eigene Verfahren zur Erkennung genannt. Die Schwierigkeit bei den genannten Verfahren liegt jedoch in der Möglichkeit zur Online-Implementierung. Daher sollen die Indizes p_5 und p_6 dies dem Anwender ermöglichen. Wie in Abschnitt 4.3 beschrieben, zeichnet sich das Regelungsszenario „Haft- und Gleitreibung“ durch typische Schwingungsformen der Regelgröße und des Stellsignals aus. Der typische Stellsignalverlauf einer Regelung mit stiction ist in Abbildung 4.20 für einen Prozess erster Ordnung mit PI-Regler, in b) für einen Prozess dritter Ordnung mit PI-Regler und in c) für einen Prozess erster Ordnung mit PID-Regler dargestellt. Es fällt auf, dass die Form im Falle eines PID-Reglers stark abweicht, daher gelten die im Folgenden eingeführten Indizes p_5 und p_6 lediglich der Diagnose von PI-geregelten Prozessen mit Ausgleich. Die meisten anderen Verfahren zur Diagnose setzen ebenfalls diese Einschränkungen voraus (dies ist [38] zu entnehmen).

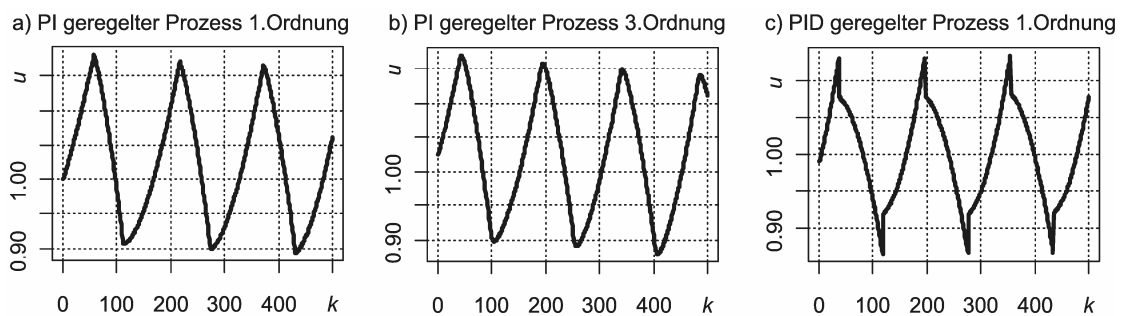


Abbildung 4.20: Stellsignalverlauf hervorgerufen durch Haft- und Gleitreibung

Anhand der simulierten PI geregelten Prozesse ist der typische dreiecksartige Verlauf des Stellsignals erkennbar. Besitzt ein Stellsignalverlauf somit eine solche Dreiecksform, so kann vermutet werden, dass das Stellgerät durch Haft- und Gleitreibung beeinflusst wird. Sowohl stochastische und deterministische Störungen als auch eine falsche Reglereinstellung rufen einen stetigeren und harmonischeren Verlauf hervor. Es wird daher zur Erkennung von Haft- und Gleitreibung die Abschätzung der Ähnlichkeit des Stellsignals zu einem Dreiecksverlauf vorgeschlagen.

Diese Abschätzung soll wie folgt geschehen: Zunächst muss das Zeitsignal u skaliert werden. Anschließend werden die Nulldurchgänge des Signals bestimmt und Zeitbereiche zwischen drei Nulldurchgängen (eine Oszillation) segmentiert. Jedes dieser Zeitsegmente muss dann separat wie folgt behandelt werden: Zunächst muss das Datensegment der Größe nach sortiert werden, wodurch der zeitliche Zusammenhang verloren geht. Anschließend werden n_K äquidistante Klassen im Bereich zwischen dem minimal skalierten und dem maximal skalierten Signalsegment definiert. Im nächsten Schritt werden die relativen Klassenhäufigkeiten h_{u_i} , basierend auf dem sortierten Signalsegment bestimmt. Trägt man Klassenmittelwerte gegenüber der relativen Häufigkeit auf, entsteht ein Histogramm. Im Falle des Dreiecksverlaufs ähnelt dieses Histogramm einer Gleichverteilung, im Falle einer Sinusschwingung einer Kamelverteilung (Namensgebung durch [27]). Dies ist in Abbildung 4.21 für eine sinusförmiges und ein dreiecksförmiges Segment gezeigt.

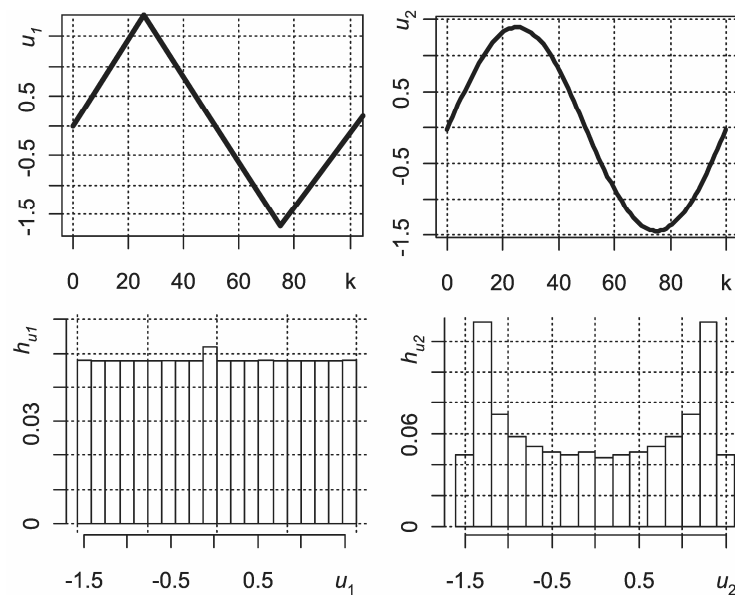


Abbildung 4.21: Histogramme (unten) für idealen dreiecksförmigen (links) und sinusförmigen Verlauf (rechts)

Die einfache Suche nach Gleichverteilung im Histogramm, etwa durch Varianzbetrachtung, reicht bei der Identifizierung des Charakteristikums nicht aus, da bereits ein geringer Rauschanteil bei dieser Art der Mustererkennung das Ergebnis verfälschen würde. Um dies zu visualisieren ist in Abbildung 4.21 derselbe Verlauf, jedoch durch weißes Rauschen gestört, dargestellt. Die Standardabweichung beträgt 5% von der Signalspannweite.

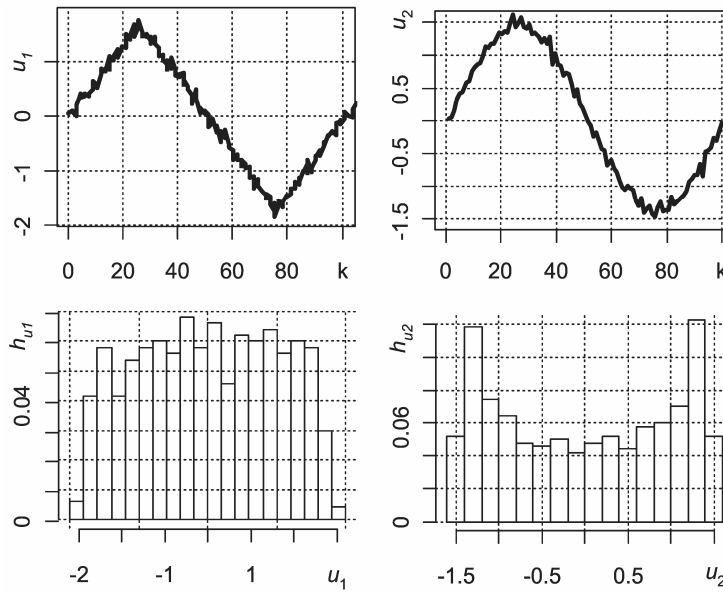


Abbildung 4.22: Histogramme (unten) für gestörten dreiecksförmigen (links) und sinusförmigen Verlauf (rechts)

Um dennoch eine Erkennung zu ermöglichen, wird die Annahme getroffen, dass alle Klassen gleichmäßig gestört sind. Dann müssen im nächsten Schritt die relativen Häufigkeiten kumulieren werden um den Effekt des weißen Rauschens zu eliminieren, wobei der Vektor h_u^{kum} entsteht. Im Falle der Sinusschwingung ähnelt der Verlauf von H_u dem einer kubischen Funktion und für den Dreiecksverlauf dem einer linearen Funktion. Die Abschätzung der Stellsignalähnlichkeit zu einem Dreiecksverlauf entspricht im letzten Schritt der Beurteilung, ob der Verlauf der kumulierten Häufigkeiten h_u^{kum} linear ist. Dies wird im Folgenden durch die Berechnung eines Hilfswerts r realisiert, welcher nach [71] definiert ist als

$$r_i = \frac{i}{n_K}, i \in [1, 2, \dots, n_K] \quad (4.12)$$

und als relativer Klassenindex anzusehen ist. Die Auftragung von h_u^{kum} über r ist für den gestörten Sinus- und Dreiecksverlauf in Abbildung 4.23 dargestellt.

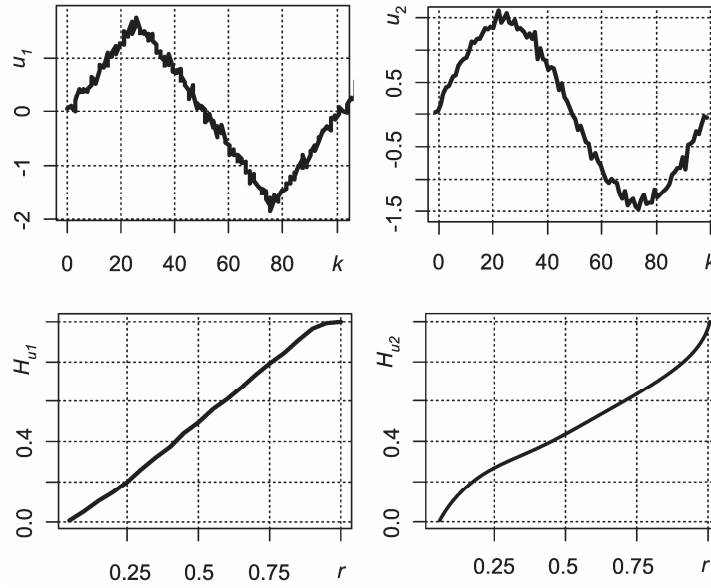


Abbildung 4.23: Kumulierte Häufigkeiten (unten) gegenüber dem relativen Klassenindex für gestörten dreiecksförmigen (links) und sinusförmigen Verlauf (rechts)

Als Maß für die Linearität der Kennlinie $h_u^{kum}(r)$ sollen die Varianzen der einzelnen Größen und die Kovarianz zwischen den beiden Größen genutzt werden. Dazu werden die Vektoren h_u^{kum} und r in n_K Paare $(h_{u,i}^{kum}, r_i)_{i=1, \dots, n_K}$ zerlegt, wobei der Wert $h_{u,i}^{kum}$ beispielsweise dem i -ten Wert im Vektor h_u^{kum} entspricht. Daraus errechnet sich die empirische Kovarianz gemäß

$$Cov(h_u^{kum}, r) = \frac{1}{n_K - 1} \sum_{i=1}^{n_K} (h_{u,i}^{kum} - \overline{h_u^{kum}}) \cdot (r_i - \bar{r}). \quad (4.13)$$

Die einzelnen empirischen Varianzen ergeben sich als

$$\begin{aligned} \sigma^2(h_u^{kum}) &= \frac{1}{n_K - 1} \sum_{i=1}^{n_K} (h_{u,i}^{kum} - \overline{h_u^{kum}})^2, \\ \sigma^2(r) &= \frac{1}{n_K - 1} \sum_{i=1}^{n_K} (r_i - \bar{r})^2. \end{aligned} \quad (4.14)$$

Die daraus ableitbare Maßzahl ε_{Eig} ist ein direktes Maß der Linearität der Kennlinie $h_u^{kum}(r)$ und somit auch ein Maß für die Dreiecksförmigkeit des zeitlichen Stellsignalverlaufs. Die Maßzahl ist definiert als

$$\varepsilon_{Eig} = \frac{a - \sqrt{a^2 - \sigma^2(h_u^{kum}) \cdot \sigma^2(r) + Cov(h_u^{kum}, r)^2}}{a + \sqrt{a^2 - \sigma^2(h_u^{kum}) \cdot \sigma^2(r) + Cov(h_u^{kum}, r)^2}}, \quad (4.15)$$

wobei $a = \frac{\sigma^2(h_u^{kum}) + \sigma^2(r)}{2}$.

Es handelt sich bei dieser Maßzahl um das Verhältnis der beiden Eigenwerte der Kovarianzmatrix von h_u^{kum} und r . Im Falle der linearen Abhängigkeit zwischen h_u^{kum} und r ist die Kovarianzmatrix singulär, besitzt somit den Eigenwert $\left(a + \sqrt{a^2 - \sigma^2(h_u^{kum}) \cdot \sigma^2(r) + Cov(h_u^{kum}, r)^2} \right)$ Null, was zu $\varepsilon_{Eig} = 0$ führt. Alternativ könnte ebenfalls die Determinante der Kovarianzmatrix als Berechnungsgröße herangezogen werden. Im Falle der Linearität und der daraus resultierenden Singularität beträgt diese Null.

Besteht keine vollständig Linearität, so weicht ε_{Eig} von Null ab. Im Folgenden soll geklärt werden, welche ε_{Eig} Werte eine signifikante Abweichung zur Linearität indizieren. Da die Grenzwerte, basierend auf Verteilungstheorien, nicht analytisch hergeleitet werden können, soll eine Monte-Carlo Simulation verwendet werden. Dafür wurden 1000 skalierte Dreiecksschwingungen simuliert, mit weißem Rauschen der Standardabweichung 0.15 addiert und die jeweiligen ε_{Eig} Werte berechnet. Die relativen Häufigkeiten von ε_{Eig} sind in Abbildung 4.24 dargestellt.

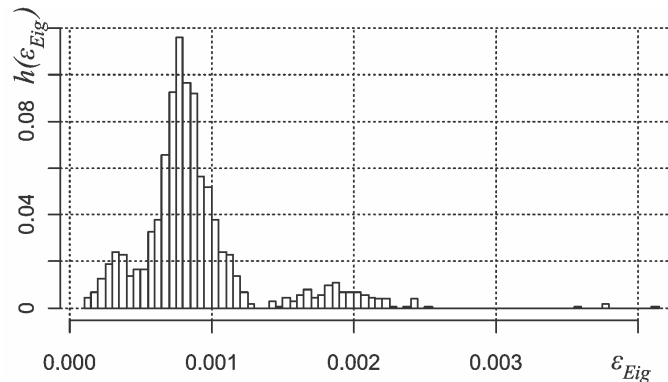


Abbildung 4.24: Relative Häufigkeiten als Ergebnis der Monte-Carlo Simulation für Dreiecksverlauf

Basierend darauf wird als Grenzwert 0.003 vorgeschlagen. Dies entspricht einer Irrtumswahrscheinlichkeit von 0.5%. Um zu zeigen, dass dieser Grenzwert geeignet gewählt wurde, soll eine zweite Monte-Carlo Simulation verwendet werden, in der im Vergleich 1000 Sinusschwingungen simuliert werden, die ebenfalls zu weißem Rauschen der Standardabweichung 0.15 addiert und die ε_{Eig} Werte berechnet wurden. Die Ergebnisse sind in Abbildung 4.25 dargestellt.

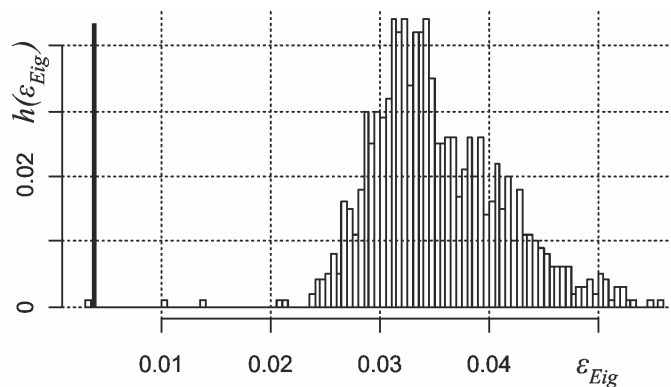


Abbildung 4.25: Relative Häufigkeiten als Ergebnis der Monte-Carlo Simulation für Sinusverlauf

Bemerkung: Die Monte-Carlo Simulation wurde mit $n_k = 20$ durchgeführt.

Aus diesen Erkenntnissen lässt sich der binäre Performance Index p_5 ableiten gemäß

$$p_5 = \begin{cases} 1, & \text{wenn } \varepsilon_{Eig} < 0.003 \\ 0, & \text{sonst} \end{cases} . \quad (4.16)$$

Dieses neue Verfahren ist jedoch nur zu verwenden, wenn die einzelnen Datensegmente aus genügend Daten bestehen. Dies bedeutet entweder eine geringe Abtastzeit oder eine genügend lange Schwingungsperiode. Der berechnete p_5 Index gilt lediglich für ein Datensegment. Der p_5 -Wert eines gesamten Datensatzes kann als Mittelwert oder Median der einzelnen p_5 -Werte bestimmt werden. Als heuristischer Grenzwert wird 0.5 verwendet. Das bedeutet, dass ein Ventil mit Haft- und Gleitreibung vermutet wird, wenn mehr als 50% der Schwingungsperioden einen eher rechteckigen Verlauf besitzen.

Position der Oszillationsmaxima, p_6

Der folgende Index basiert auf der Idee von [54] und nutzt die Position der Oszillationsmaxima des Regelfehlers zur Identifikation von Stiction. Für den Fall, dass die Regelgröße schwingt und der Grund hierfür eine externe Störung oder aber eine aggressive Reglereinstellung ist, befindet sich das Maximum zwischen zwei Nulldurchgängen des Regelfehlers idealerweise in der Mitte. Im Falle einer, durch Stiction hervorgerufenen Schwingung ist dies nicht der Fall. Die Position des Maximums ist, abhängig vom Prozess, nach links oder rechts verschoben. Sei $k_{0\max,i}$ die Position des Maximums des absoluten Regelfehlers zwischen zwei Nulldurchgängen und $\Delta k_{0,i}$ die Anzahl an Abtastpunkten zwischen dem i -ten und dem $(i+1)$ -ten Nulldurchgang, so lässt sich für jede Halbperiode die relative Position des i -ten Maximums bestimmen als

$$\theta_i = k_{0\max,i} / \Delta k_{0,i} . \quad (4.17)$$

Daraus lässt sich p_6 somit als ein Maß für die Ausprägung von Stiction in einem Prozess mit Ausgleich definieren gemäß

$$p_6 = 2 \cdot \left| 0.5 - \frac{1}{l-1} \sum_{i=1}^{l-1} \theta_i \right| . \quad (4.18)$$

Es gilt, dass $p_6 \in [0,1]$, wobei $p_6 = 0$ einem mittigen Maximum entspricht. Um einen Grenzwert festlegen zu können, werden in einer Monte-Carlo-Simulation die p_6 -Werte für 1000 Datensätze berechnet, die ihrerseits aus 20 sinusförmigen Halbschwingungen ($l = 20$) bestehen. Jeder Datensatz wurde mit weißem Rauschen gestört, dessen Standardabweichung 5% der Spannweite beträgt. Die Verteilung der 1000 berechneten p_6 -Werte ist in Abbildung 4.26 dargestellt. Als Grenzwert wird, basierend auf der Verteilung, ein Grenzwert von 0.15 gewählt. Dies hätte bei der Monte-Carlo-Simulation zu 2% Fehlentscheidungen geführt. Der Grenzwert ist daher so gering gewählt, da die Position des Maximums sich auch bei Stictionschwingungen zum Teil nur gering von der Mitte abweicht. Außerdem sollte der Index nur berechnet werden, wenn $\Delta k_{0,i}$ einen

gewissen Wert überschreitet (in der Praxis hat sich hierfür ~ 5 als gut erwiesen). Kleinere Werte und somit dichtere Nulldurchgänge werden in aller Regel nur durch weißes Rauschen hervorgerufen. Wird dieser Wert (~ 5) nicht erreicht, so ist θ_i als 0.5 anzunehmen und es ergibt sich $p_6 = 0$.

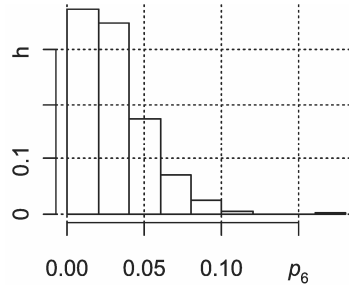


Abbildung 4.26: Monte-Carlo basierte Verteilung der berechneten p_6 -Werte

Normalisierter Oszillationsindex des Regelfehlers, p_7

Um Auskunft über die Periodizität des Regelfehlers geben zu können, wird p_7 für die Autokorrelationsfunktion des Regelfehlers acf_e gemäß (3.15) nach [71] definiert als

$$p_7 = n_{oi}(acf_e). \quad (4.19)$$

Basierend auf den bisherigen Erkenntnissen wird als kritischer Grenzwert 0.71 gewählt.

Normalisierter Oszillationsindex des Sollwerts, p_8

Die Periodizität des Sollwerts wird durch p_8 bestimmt, welcher für die Autokorrelationsfunktion des Sollwerts acf_w nach [71] definiert ist als

$$p_8 = n_{oi}(acf_w). \quad (4.20)$$

Auch für die Schwingung des Sollwerts als kritischer Grenzwert 0.71 gewählt.

Beeinflussung durch deterministische externe Störungen, p_9

Zur Abschätzung des Einflusses von deterministischen, externen Störungen soll p_9 dienen. Es wird dabei die Eigenschaft der Störungsübertragungsfunktion genutzt, welche sich bei der Bedingung des konstanten Sollwerts (o.B.d.A $w = 0$), basierend auf der Darstellung eines geschlossenen Regelkreises aus Abbildung 3.1, mit der allgemeinen Übertragungsfunktion des Reglers $G_C(s)$ und der Übertragungsfunktion des Prozesses $G_P(s)$ berechnet zu

$$Y(s) = \frac{1}{1 + G_c(s)G_p(s)e^{-sT_d}} E(s). \quad (4.21)$$

Das bedeutet, dass im Gegensatz zu einer Sollwertänderung, eine Störgrößenänderung (nach obigem Modell) eine sofortige Auswirkung auf die Regelgröße besitzt. Dies gilt auch für den Fall, dass der Regler einen differenzierenden Anteil besitzt, da die dadurch hervorgerufenen abrupten Änderungen durch den Prozess

(der in der Praxis nicht ideal differenzierend ist) verzögert werden. Um die Beeinflussung durch externe Störungen abschätzen zu können, wird somit nach schnellen Veränderungen der Regelgröße gesucht.

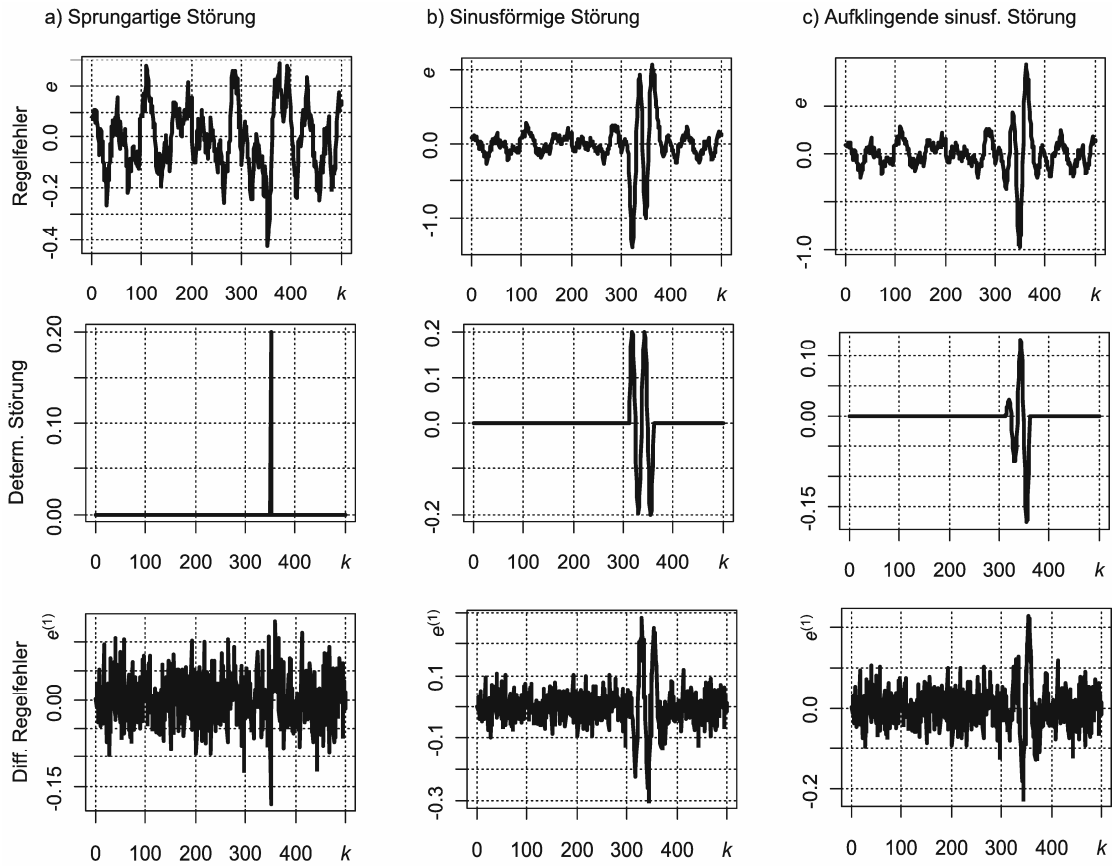


Abbildung 4.27: Verschiedene deterministische Störungen und deren Einfluss auf den Regelfehler

Bildet man die erste diskrete Differenzierung der zu analysierenden Messreihe e ($e_k - e_{k-1}$), so treten solche Veränderungen noch deutlicher hervor. Um dies zu verdeutlichen sind in Abbildung 4.27 verschiedene deterministische Störungen für einen Prozess

erster Ordnung mit Totzeit $\left(G_p(s) = \frac{1}{s+1} \cdot e^{-s \cdot 0.25}, \Delta T = 0.05 \right)$ simuliert mit einem PI-

Regler ($K_{PR} = 1.5$, $T_N = 0.7$). Die deterministischen Störungen sind dabei in a) eine sprunghafte, in b) eine kurzzeitige Sinusschwingung und in c) eine langsam ansteigende Sinusschwingung. Zusätzlich zu diesen deterministischen Störungen wird weiterhin eine stochastische Störung addiert (Mittelwert = 0, Varianz = 0.0016). Die Abschätzung von deterministischen Störungen kann durch einen klassischen Ausreißertest, basierend auf dem differenzierten Regelfehler geschehen. Die beste Art hierfür bietet der robuste Ausreißertest, dessen Ausreißergrenzen auf dem Median einer Messreihe basieren. Verschiedene Verfahren (z.B. [50]) nutzen robuste Kenngrößen zur Schätzung der Ausreißergrenzen, da bei dieser Art die Grenzen selbst kaum von den Ausreißern beeinflusst werden. Die boolsche Variable ε_{out} zur Ausreißererkennung wird definiert gemäß

$$\varepsilon_{out,k} = \begin{cases} 1, & \text{wenn } |e_k^{(1)} - e_{0.5}^{(1)}| > 3 \cdot 1.4826 \cdot MAD(e^{(1)}) \\ 0, & \text{sonst} \end{cases}, \quad (4.22)$$

wobei $e_{0.5}^{(1)}$ dem Median des differenzierten Regelfehlers und $MAD(\cdot)$ dem Median der absoluten Medianabweichung entspricht.

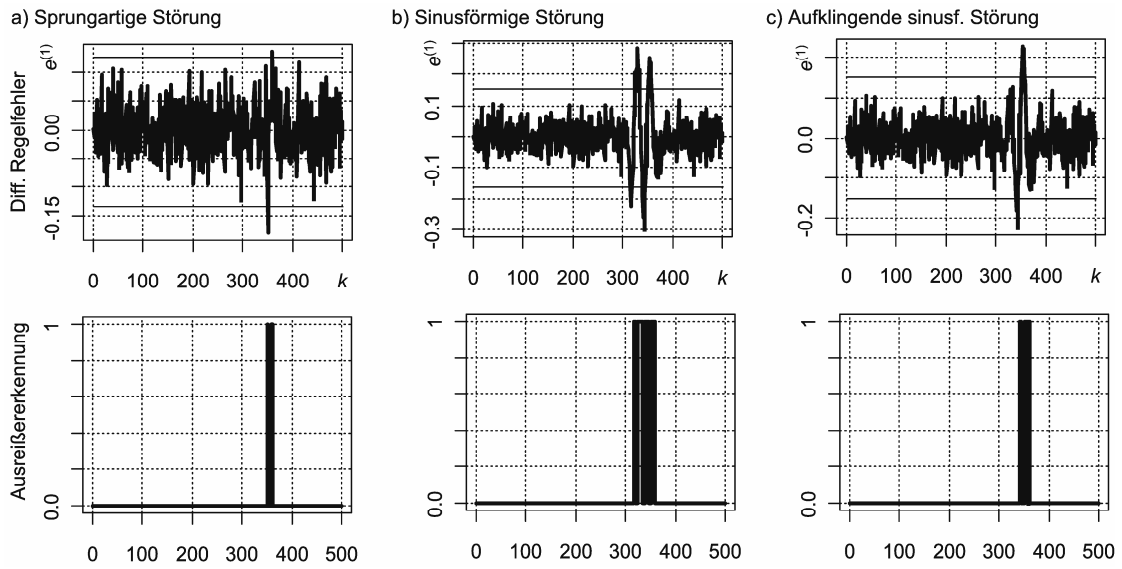


Abbildung 4.28: Verschiedene deterministische Störungen mit angewandter Ausreißererkennung

Wird angenommen, dass $e^{(1)}$ annähernd normal verteilt ist, so entspricht $e_{0.5}^{(1)}$ der robusten Schätzung des Mittelwerts und $1.4826 \cdot MAD(e^{(1)})$ dem der Standardabweichung. 99.5% aller normal verteilten $e^{(1)}$ -Daten liegen innerhalb des Bereichs von $e_{0.5}^{(1)} \pm 3 \cdot 1.4826 \cdot MAD(e^{(1)})$. Die Grenzen sind in Abbildung 4.28 für die Simulationen aus Abbildung 4.27 dargestellt. Die boolsche Variable zur Ausreißererkennung ist ebenfalls dargestellt. Basierend auf der Variable ε_{out} wird der Performance Index zur Abschätzung der Beeinflussung durch deterministische Störungen definiert als

$$p_9 = \frac{1}{n} \sum_{i=1}^n \varepsilon_{out,i}. \quad (4.23)$$

Die Verwendung des vorgestellten Index p_9 ist nur anzuwenden, wenn die Bedingungen erfüllt sind, dass der Sollwert sich nicht oder nur sehr langsam ändert. Oder aber bei sich schnell änderndem Sollwert ist vorauszusetzen, dass der Prozess in Kombination mit dem Regler genügend träge ist. Weitere Einschränkungen liegen darin, dass die deterministische Störung zwischen ihrer Entstehung und der Wirkung auf den Prozess keine Dynamik durchlaufen darf, die träger ist, als die des Prozesses. Außerdem muss die Anzahl der Störungen gering im Bezug auf die Anzahl an Messungen sein, da die Grenzen der Ausreißererkennung sonst trotz robuster Schätzung verfälscht würden. Simulationen zeigen, dass externe Störungen während 10% bis 20% der Datenaufzeichnung als Maximalwert angesehen werden können.

Basierend auf der Wahl der Ausreißergrenzen $e_{0.5}^{(1)} \pm 3 \cdot 1.4826 \cdot MAD(e^{(1)})$ liegen circa 99.5% aller $e_k^{(1)}$ Daten innerhalb dieser. Das bedeutet theoretisch, dass p_9 -Werte ab 0.005 ein Anzeichen für externe Störungen sind. Da diese Begründung jedoch nur unter der Annahme der Normalverteilung erfüllt ist, soll als Grenze das Zehnfache des theoretischen Werts, also 0.05 genommen werden.

Normalisierter Idle-Index, p_{10}

Die Erkennung einer langsamen Regelung ist nach [24] durch die Berechnung des so genannten Idle-Index möglich. Bedingt durch die Tatsache, dass die träge Reglereinstellung nur geringe Auswirkung auf die anlagenweite Regelungsperformance hat und die Tatsache, dass die Berechnung des Idle-Index die große Einschränkung besitzt, dass er nur nach signifikanten Sollwertsprüngen berechenbar ist, soll er nur der Vollständigkeit halber erwähnt werden. Er spielt bei späteren Berechnungen keine Rolle. Andere Verfahren zur Abschätzung der Reglerträgheit, wie [32] nutzen dazu die Mustererkennung und setzen vorherige Benchmarkprozesse voraus, sind somit ebenfalls nur begrenzt automatisierbar.

Wie in Abschnitt 4.2 erwähnt, zeichnet sich eine langsame Regelung durch die Gleichgerichtetheit zwischen Stellsignal und Regelgröße aus. Seien u_{ins} und y_{ins} die durch einen Sollwertsprung hervorgerufenen instationären Bereiche des vorliegenden Datensatzes, so schlägt [24] zunächst die Bestimmung der Zeit gleichsinniger t_{pos} und gegensinniger Bewegung t_{neg} vor, nach

$$\begin{aligned} t_{pos} &= \begin{cases} t_{pos} + \Delta T, & \text{wenn } \Delta u_{ins} \cdot \Delta y_{ins} > 0 \\ t_{pos}, & \text{wenn } \Delta u_{ins} \cdot \Delta y_{ins} \leq 0 \end{cases} \\ t_{neg} &= \begin{cases} t_{neg} + \Delta T, & \text{wenn } \Delta u_{ins} \cdot \Delta y_{ins} < 0 \\ t_{neg}, & \text{wenn } \Delta u_{ins} \cdot \Delta y_{ins} \geq 0 \end{cases}. \end{aligned} \quad (4.24)$$

Daraus berechnet sich bei positiver Prozessverstärkung der Idle Index i^+ nach

$$i^+ = \frac{t_{pos} - t_{neg}}{t_{pos} + t_{neg}}. \quad (4.25)$$

Bei einer negativer Prozessverstärkung berechnet er sich zu

$$i^- = \frac{-t_{pos} + t_{neg}}{t_{pos} + t_{neg}}. \quad (4.26)$$

Es gilt die Normierung $-1 \leq i^-, i^+ \leq 1$. Nach [24] gilt ein Regler ab einem Idle-Index von 0.5 als träge. Kleinere Werte können jedoch nicht als Anzeichen einer aggressiven Regelung gedeutet werden. Basierend auf dem Wunsch, dass alle Performance-Indizes zwischen 0 und 1 liegen sollen, lässt sich die lineare Transformation ableiten gemäß

$$p_{10} = \frac{1}{2} \left(\frac{\text{sign}(K_{PS}) \cdot t_{pos} - \text{sign}(K_{PS}) \cdot t_{neg}}{t_{pos} + t_{neg}} + 1 \right). \quad (4.27)$$

Basierend auf der Transformation sind p_{10} -Werte ab 0.75 als Anzeichen einer trägen Regelung anzusehen.

4.5 Zusammenfassende Bemerkung

In Kapitel 4 wurden die gängigsten Regelungsszenarien und das Szenario der Haft- und Gleitreibung in Ventilkörpern vorgestellt und deren Charakteristika genannt. Darauf basierend wurden in 4.4 zehn Performance-Indizes entweder entwickelt oder geeignete Herangehensweise der Literatur entnommen, welche eine automatisierbare Diagnose ermöglichen. Eine Ausnahme hierfür ist der normierte Idle-Index, der nur aus Gründen der Vollständigkeit erwähnt wurde. Zusammengefasst ergibt sich, basierend auf der Charakteristikadarstellung aus Tabelle 4.3 (erweitert aus [71]) der Zusammenhang der zehn Performance-Indizes und der einzelnen Regelungsszenarien aus Tabelle 4.2. Das Symbol \uparrow bei einem Performance-Index p_i gibt an, dass der p_i -Wert für das entsprechende Szenario groß (größer als genannter Grenzwert) sein muss. Das Symbol \downarrow gibt an, dass der Index möglichst klein (kleiner als der genannte Grenzwert) sein muss und das Symbol \rightarrow gibt an, dass der entsprechende Index keine Rolle bei der Beurteilung des genannten Regelungsszenarios spielt.

Tabelle 4.3: Zusammenhang zwischen Regelungssituation und Performance-Indizes

	p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8	p_9	p_{10}
Aggressive Reglereinstellung	\downarrow	\rightarrow	\rightarrow	\rightarrow	\downarrow	\downarrow	\uparrow	\downarrow	\rightarrow	\rightarrow
Schwingender Sollwert	\rightarrow	\rightarrow	\rightarrow	\rightarrow	\rightarrow	\rightarrow	\rightarrow	\uparrow	\rightarrow	\rightarrow
Messrauschen	\rightarrow	\rightarrow	\rightarrow	\uparrow	\rightarrow	\rightarrow	\rightarrow	\rightarrow	\rightarrow	\rightarrow
Begrenzungen	\uparrow	\uparrow	\uparrow	\rightarrow	\rightarrow	\rightarrow	\rightarrow	\rightarrow	\rightarrow	\rightarrow
Inaktive Regelung	\uparrow	\uparrow	\downarrow	\rightarrow	\rightarrow	\rightarrow	\rightarrow	\rightarrow	\rightarrow	\rightarrow
Träge Reglereinstellung	\rightarrow	\rightarrow	\rightarrow	\rightarrow	\rightarrow	\rightarrow	\downarrow	\downarrow	\rightarrow	\uparrow
Haft- und Gleitreibung	\rightarrow	\rightarrow	\rightarrow	\rightarrow	\uparrow	\uparrow	\uparrow	\downarrow	\rightarrow	\rightarrow
Externe Störungen	\rightarrow	\rightarrow	\rightarrow	\rightarrow	\rightarrow	\rightarrow	\rightarrow	\rightarrow	\uparrow	\rightarrow

Um die Wirksamkeit des vorgestellten Diagnoseverfahrens zu zeigen, seien erneut die Benchmarkszenarien aus Abbildung 3.2 aufgegriffen (nahezu optimale Regelung mit 20% Überschwung, aggressive Regelung, Stiction und begrenzte Regelung).

Es ergeben sich die berechneten Werte aus Tabelle 4.4. Alle Signale sind dabei zur Quantisierung auf 4 Nachkommastellen gerundet.

Tabelle 4.4: Berechnete Performance-Indizes der Benchmarkprozesse

	p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8	p_9
Optimale Regelung	0.010	0	0	0.045	0	0	0.579	0	0.042
Aggressive Regelung	0.005	0	0.052	0.034	0	0.062	0.766	0	0.083
Haft- und Gleitreibung	0	0	0	0.027	1	0.284	0.981	0	0.038
Begrenzungen	0.997	0.75	1	0.027	0	0	0	0	0.057

Bezieht man die genannten Grenzwerte ein und leitet daraus die grafische Darstellung \downarrow (kleiner als Grenzwert) und \uparrow (größer als Grenzwert) ab, so ergibt sich die Darstellung aus Tabelle 4.5. Hierbei sind die Darstellungen, basierend auf den Messungen, den Darstellungen, basierend auf idealen Werten, gegenübergestellt. Eine einfache Klassifizierung erkennt alle Regelungsszenarien anhand der Charakteristikaidentifikation einwandfrei.

Tabelle 4.5: Berechnete Performance-Indizes der Benchmarkprozesse

	p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8	p_9
Optimale Regelung	\downarrow	\downarrow	\downarrow	\downarrow	\downarrow	\downarrow	\downarrow	\downarrow	\downarrow
Aggressive Regelung (Messung)	\downarrow	\downarrow	\downarrow	\downarrow	\downarrow	\downarrow	\uparrow	\downarrow	\downarrow
Aggressive Reglereinstellung (ideal)	\downarrow	\rightarrow	\rightarrow	\rightarrow	\downarrow	\downarrow	\uparrow	\downarrow	\uparrow
Haft- und Gleitreibung (Messung)	\downarrow	\downarrow	\downarrow	\downarrow	\uparrow	\uparrow	\uparrow	\downarrow	\downarrow
Haft- und Gleitreibung (ideal)	\rightarrow	\rightarrow	\rightarrow	\rightarrow	\uparrow	\uparrow	\uparrow	\downarrow	\rightarrow
Begrenzungen (Messung)	\uparrow	\uparrow	\uparrow	\downarrow	\downarrow	\downarrow	\downarrow	\downarrow	\uparrow
Begrenzungen (ideal)	\uparrow	\uparrow	\uparrow	\rightarrow	\rightarrow	\rightarrow	\rightarrow	\rightarrow	\rightarrow

Neben dieser einfachsten Art der Klassifizierung gibt es ebenfalls andere Ansätze, wie beispielsweise die der Klassifikation durch neuronale Netze, Bayes Wahrscheinlichkeiten [43] oder Support Vector Maschinen [56]. Bei Verwendung solcher komplexer Herangehensweisen steigt zwar der Rechenbedarf aber es sind z.B. Aussagen über Zugehörigkeitswahrscheinlichkeiten möglich, wie z.B. Datensatz X zeigt mit 93 prozentiger Wahrscheinlichkeit eine aggressive Regelung. Außerdem können die genannten Grenzwerte durch die Auswertung historischer Betriebsdaten angepasst

werden, was das Ergebnis der Beurteilung verbessert aber auch die Ansprüche an das Wissen des Personals heraufsetzt.

Kapitel 5

Validierung und Testen an Benchmark-Anlagen

Dieses Kapitel beschäftigt sich mit der Validierung und dem Testen der genannten Verfahren der anlagenweiten Regelkreisdiagnostik an simulierten und industriellen Benchmark-Anlagen. Es ist in zwei Abschnitte aufgeteilt, wobei in Abschnitt 5.1 die Validierung anhand einer simulierten Benchmark-Anlage stattfindet. Dazu werden verschiedene Regelungsszenarien (siehe Abschnitt 1.1) in einzelnen Regelkreisen simuliert, welche zu anlagenweiten Fehlern führen. Mittels der in Kapitel 2 genannten Verfahren wird zunächst versucht den Verursacher zu identifizieren und anschließend mittels der Fehlerdiagnose aus Kapitel 4 wird versucht das simulierte Szenario korrekt zu diagnostizieren. In Abschnitt 5.2 werden die so validierten Verfahren genutzt um eine industrielle Anlage zu diagnostizieren. Auch die industriellen Daten zeigen anlagenweite Fehler, jedoch mit unbekannter Ursache.

5.1 Simulierte Benchmarks

Die Validierung an simulierten Benchmarks geschieht an einer simulierten Hydrocrackeranlage.

Hydrocrackeranlage - Verfahrensbeschreibung:

Hydrocracker werden eingesetzt, um schwer siedende Kohlenwasserstoffe in wirtschaftlich nutzbarere Kohlenwasserstoffe aufzuspalten, welche einen geringeren Siedepunkt haben, wie z.B. Naphta, Kerosin oder Diesel. Dies geschieht unter hohem Druck durch die Zugabe von Wasserstoff. Die Reaktion findet dabei in einer Reihe von aufeinander folgenden katalytischen Festbetten statt. Der verwendete Wasserstoff wird neben der eigentlichen stöchiometrischen Notwendigkeit auch zur Temperaturregelung in den so genannten Quenchen (statische Mischer) eingesetzt. Eine solche Temperaturregelung ist notwendig, da die Reaktion der Aufspaltung stark exotherm ist und, wie die meisten chemischen Reaktionen, je schneller verläuft, desto wärmer die umgebende Temperatur ist. Die Simulation der Hydrocrackeranlage basiert auf der Modellidentifikation einer industriellen Hydrocrackeranlage. Das vereinfachte Fließbild dieser Anlage ist in Abbildung 5.1 zu sehen.

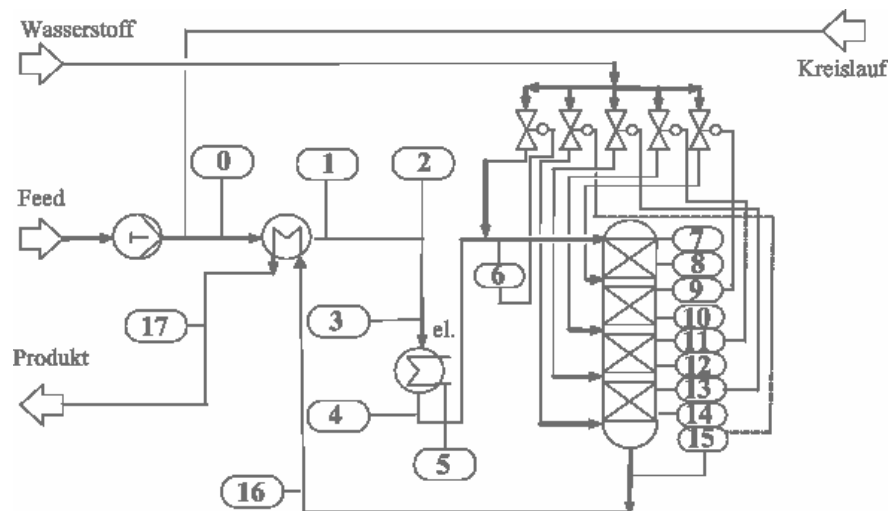


Abbildung 5.1: Vereinfachtes Verfahrensfliessbild einer industriellen Hydrocrackeranlage

Die Simulation wurde dabei unter Verwendung von Matlab Simulink[®] durchgeführt. Das gesamte Simulationsmodell ist in Abbildung 5.2 gezeigt.

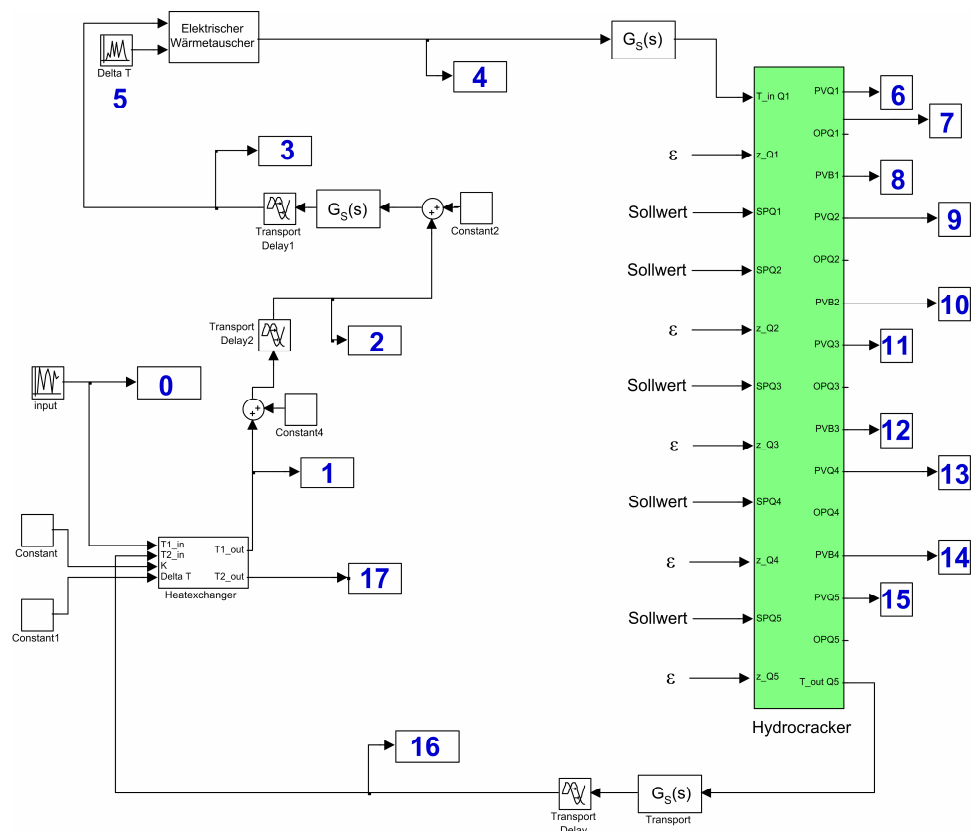


Abbildung 5.2: Modell einer simulierten Hydrocrackeranlage

Das Simulationsmodell besteht, genau wie die industrielle Anlage, aus vier katalytischen Festbetten und fünf Quenches zur Temperaturregelung. Deren Modelle sind in Abbildung 5.3 und Abbildung 5.4 gezeigt. Außerdem bietet die Simulation die Möglichkeit externe Störungen in jedem Quench zu generieren.

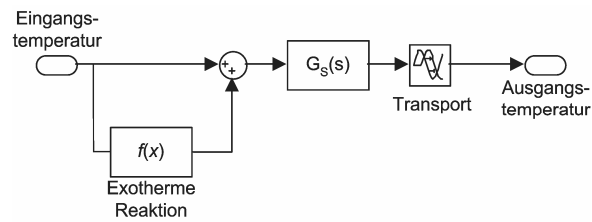


Abbildung 5.3: Modell eines simulierten katalytischen Festbetts mit aufgeschalteter exothermer Temperaturzunahme $f(x)$

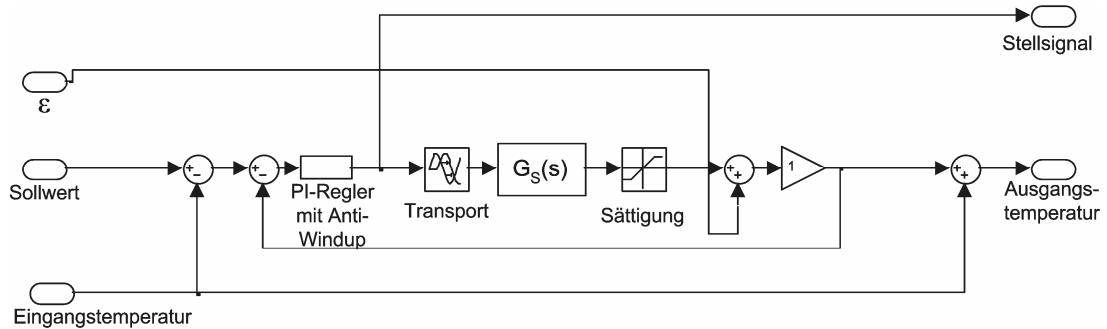


Abbildung 5.4: Modell einer simulierten Quenchs

Mithilfe des Modells sollen die folgenden Regelungsszenarien simuliert werden:

- Stiction in S_6 (Messstelle Nr. 6)
- Oszillierender Sollwert S_{11} (Messung Nr. 11)

Stiction in S_6 :

Für die erste Validierungssimulation wurde Haft- und Gleitreibung im ersten Quench simuliert. Die simulierten Messdaten aller Messstellen sind in Abbildung 5.5 dargestellt.

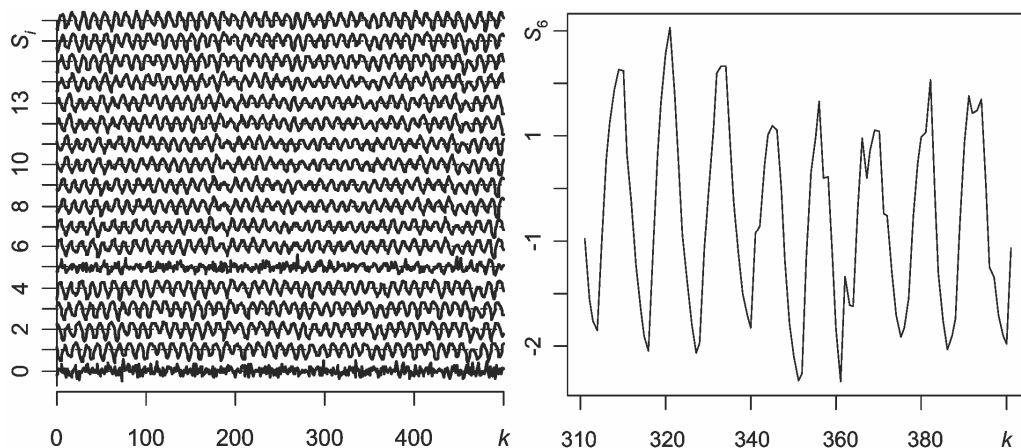


Abbildung 5.5: Daten aller Messungen (links) und simulierter Haft- und Gleitreibung an Messung Nr.6 (S_6) (rechts)

Basierend auf der in Abbildung 2.2 dargestellten Prozedur muss zunächst der Verursacher identifiziert werden. Dies kann, je nach verfügbarer Rechenleistung, entweder nach 2.2 (Offline-Identifizierung) oder aber mithilfe des stark vereinfachten Verfahrens nach Abschnitt 2.3 (Online-Identifizierung) durchgeführt werden. Im Folgenden sollen beide Verfahren getestet werden.

Offline-Verursacheridentifizierung mittels KNNK Algorithmus gemäß Abschnitt 2.2

Es ergibt sich die folgende Beeinflussungsmatrix für die Daten aus Abbildung 5.5, wobei zu beachten ist, dass Messstelle 0 als Anlageneintritt S_0 betrachtet werden kann, und somit bei der Berechnung der Beeinflussungsmatrix vernachlässigt wurde.

$$\mathbf{Z} = \begin{bmatrix} - & & & & & & & & & & & & & & & & & & \\ 1 & - & & & & & & & & & & & & & & & & & \\ 1 & 2 & - & & & & & & & & & & & & & & & & \\ 1 & 2 & 3 & - & & & & & & & & & & & & & & & \\ 1 & 2 & 3 & 4 & - & & & & & & & & & & & & & \\ 6 & 6 & 6 & 6 & 6 & - & & & & & & & & & & & & \\ 7 & 7 & 7 & 7 & 5 & 6 & - & & & & & & & & & & & \\ 8 & 8 & 8 & 8 & 5 & 6 & 7 & - & & & & & & & & & & \\ 9 & 9 & 9 & 9 & 5 & 6 & 7 & 8 & - & & & & & & & & & \\ 10 & 10 & 10 & 10 & 5 & 6 & 7 & 8 & 9 & - & & & & & & & & \\ 11 & 11 & 11 & 11 & 5 & 6 & 7 & 8 & 9 & 10 & - & & & & & & & \\ 12 & 12 & 12 & 12 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & - & & & & & & \\ 13 & 13 & 13 & 13 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & - & & & & & \\ 14 & 14 & 14 & 14 & 14 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & - & & & & \\ 15 & 15 & 15 & 15 & 15 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & - & & & \\ 16 & 16 & 16 & 16 & 16 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 & - & & \\ 17 & 17 & 17 & 17 & 17 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 & 16 & - & \end{bmatrix}. \quad (5.1)$$

Der Verursacher wurde somit, basierend auf der Analyse der Beeinflussungsrichtung, korrekt identifiziert (erkennbar an der 6-ten Zeile und Spalte, die nur die Einträge „6“ aufweist).

Online-Verursacheridentifizierung anhand der Störungsenergie gemäß Abschnitt 2.3

Die vereinfachte Online-Verursacheridentifikation basierend anhand des Energiespektrums eignet sich nur begrenzt für das Regelungsszenario „Haft- und Gleitreibung“, da in diesem Fall die hervorgerufene Oszillation nicht einer idealen Sinusschwingung gleicht. Dies hat wiederum zur Folge, dass die Störungsenergie nicht alleine in einem zusammenhängenden Störungsfrequenzband (wie es für diese Vereinfachung vorausgesetzt wird) übertragen wird, sondern in äquidistanten Frequenzbändern. Hervorgerufen wird dies durch die rechteckige Schwingungsform. In Abbildung 5.6 ist das Spektrum des Verursachers S_6 dargestellt, die Fundamentale und ihre ersten beiden Harmonischen sind ebenfalls gekennzeichnet.

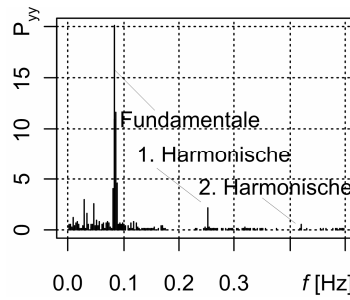
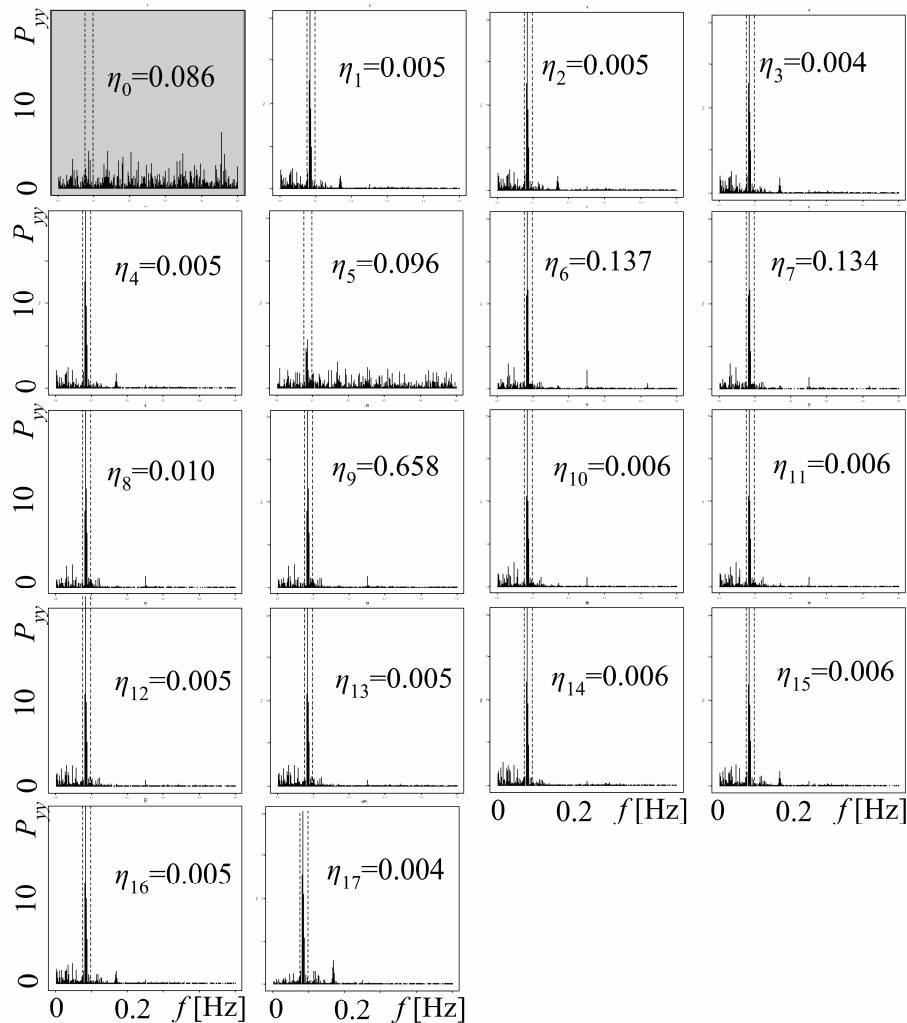


Abbildung 5.6: Spektrum des verursachenden Regelkreises

Abbildung 5.7: Spektren aller Teilprozesse bei Verursacher S_6 (grau unterlegt: ausgeschlossen von Analyse, da dort die anlagenweite Störfrequenz nicht auffindbar ist)

Für dieses Regelungsszenario müsste somit die relative übertragene Störungsenergie anhand von mehreren Störfrequenzbändern berechnet werden. Dies würde zu einem idealen Ergebnis führen, da nach [60] die harmonischen Anteile der Schwingung durch die Teilprozesse gefiltert werden. Da die Identifikation der Fundamentalen und der Harmonischen sehr schwierig zu automatisieren ist und dies somit den Voraussetzungen dieses Online-Verfahrens widerspricht, sollen die relativen Störungsenergien lediglich basierend auf der Fundamentalen berechnet werden, da angenommen wird, dass diese im Vergleich zu den Harmonischen den größten Anteil an der hervorgerufenen Störung besitzt. Vernachlässigt man den Anlageneintritt S_0 so lässt sich das anlagenweite

Störungsfrequenzband (Fundamentale) zwischen 0.074 und 0.098 Hz identifizieren. Darauf basierend berechnen sich sämtliche relative Störungsenergien gemäß Abbildung 5.7. An der Abbildung ist ebenfalls gut zu erkennen, dass die erste Harmonische bei ~0.25 Hz tatsächlich auf ihrem Weg durch die Anlage abgeschwächt wird. Eine weitere Schwingung bei ~0.18 Hz wird durch die Regelung im letzten Quench (Messung 15) hervorgerufen.

Detaillierte Fehlerdiagnose im verursachenden Teilprozess S_6

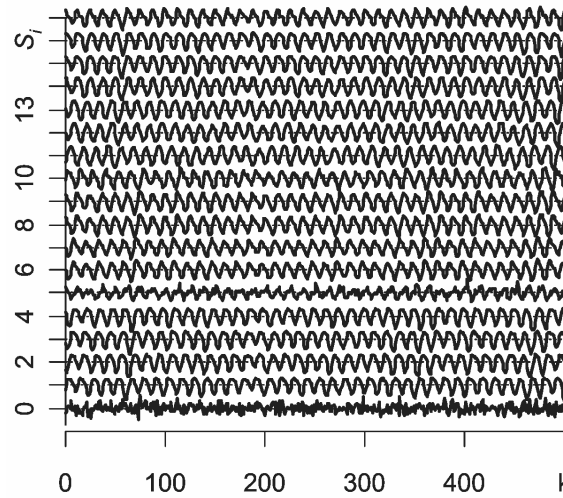
Basierend auf den Ergebnissen der erfolgreichen Verursacheridentifikation beider Verfahren und der in Abbildung 2.2 vorgeschlagenen Prozedur muss im nächsten Schritt eine Fehlerdiagnose im Teilprozess S_6 durchgeführt werden. Dazu werden die in Kapitel 4 vorgestellten Verfahren verwendet. Es werden die Performance-Indizes gemäß Tabelle 5.1 berechnet. Das Regelungsszenario Haft- und Gleitreibung wird somit korrekt diagnostiziert. Außerdem gibt es Anzeichen für starkes Messrauschen.

Tabelle 5.1: Berechnete Performance-Indizes des Benchmarkprozesses bei Haft- und Gleitreibung in S_6
(grau hinterlegt: Anforderung erfüllt, schraffiert: Anforderung nicht erfüllt)

	p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8	p_9
S_7	0.002 ↓	0 ↓	0 ↓	0.146 ↑	0.844 ↑	0.233 ↑	0.972 ↑	0 ↓	0.002 ↓
Aggressive Reglereinstellung	↓	→	→	→	↓	↓	↑	↓	→
Schwingender Sollwert	→	→	→	→	→	→	→	↑	→
Messrauschen	→	→	→	↑	→	→	→	→	→
Begrenzungen	↑	↑	↑	→	→	→	→	→	→
Inaktive Regelung	↑	↑	↓	→	→	→	→	→	→
Haft- und Gleitreibung	→	→	→	→	↑	↑	↑	↓	→
Externe Störungen	→	→	→	→	→	→	→	→	↑

Oszillierender Sollwert in S_{11} :

Für die folgende Simulation wurde ein sinusförmig schwingender Sollwert für Messung 11 (S_{11}) angenommen, wie er etwa bei einer aggressiven überlagerten Kaskadenregelung auftreten kann. Die Periodendauer der Schwingung beträgt 12.5 Abtastzeiten mit dem Mittelwert 410°C und einer Amplitude von 10°C. Die Aufzeichnungen aller Messungen sind in Abbildung 5.8 dargestellt.

Abbildung 5.8: Daten aller Messungen bei oszillierendem Sollwert in S_{11} [74]

Offline-Verursacheridentifizierung mittels KNNK Algorithmus gemäß Abschnitt 2.2

Es ergibt sich die Beeinflussungsmatrix für die Daten aus Abbildung 5.8 gemäß

$$\mathbf{Z} = \begin{bmatrix} - & & & & & & & & & & & & & & & & & & \\ 1 & - & & & & & & & & & & & & & & & & & \\ 1 & 2 & - & & & & & & & & & & & & & & & & \\ 1 & 2 & 3 & - & & & & & & & & & & & & & & & \\ 1 & 2 & 5 & 5 & - & & & & & & & & & & & & & \\ 1 & 2 & 3 & 6 & 5 & - & & & & & & & & & & & & \\ 1 & 2 & 3 & 7 & 5 & 6 & - & & & & & & & & & & & \\ 1 & 2 & 3 & 4 & 8 & 6 & 7 & - & & & & & & & & & & \\ 1 & 2 & 3 & 4 & 9 & 6 & 7 & \phi & - & & & & & & & & & \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & - & & & & & & & & \\ 11 & 11 & 11 & 11 & 11 & 11 & 11 & 11 & 11 & 11 & - & & & & & & & \\ 12 & 2 & 3 & 12 & 5 & 12 & 12 & 8 & 9 & 12 & 11 & - & & & & & & \\ 13 & 2 & 3 & 13 & 5 & 13 & 13 & 8 & 9 & 13 & 11 & 12 & - & & & & & \\ 14 & 14 & 14 & 14 & 14 & 14 & 14 & \phi & \phi & 10 & 11 & 12 & 13 & - & & & & \\ 15 & 15 & 15 & 15 & 15 & 15 & 15 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & - & & & \\ 16 & 16 & 16 & 16 & 16 & 16 & 16 & 8 & 9 & 16 & 11 & 12 & 13 & 14 & 15 & - & & \\ 17 & 17 & 17 & 17 & 17 & 17 & 17 & 8 & 9 & 17 & 11 & 12 & 13 & 14 & 15 & 16 & - & \end{bmatrix}, \quad (5.2)$$

wobei zu beachten ist, dass Messstelle 0 als Anlageneintritt S_0 betrachtet werden kann, und somit bei der Berechnung der Beeinflussungsmatrix vernachlässigt wurde.

Der Verursacher wurde somit, basierend auf der Analyse der Beeinflussungsrichtung, korrekt identifiziert. Die Kennzeichnung ϕ zeigt an, dass bei dieser Berechnung keine Entscheidung bzgl. des Verursachers möglich war. Ein Grund hierfür kann sein, dass es keinen eindeutigen Maximalwert der Funktion $\rho(n_d)$ gibt.

Online-Verursacheridentifizierung anhand der Störungsenergie gemäß Abschnitt 2.3

Die vereinfachte Verursacheridentifikation eignet sich in diesem Fall ausgezeichnet für die Verursacheridentifikation, da es sich um ein sinusförmiges Verursachersignal handelt. Die Spektren aller Messung sind in Abbildung 5.9 dargestellt. Wird die Messung 0 vernachlässigt, so ergibt sich das Störungsfrequenzband im Bereich von 0.058 bis 0.098 Hz. Der verursachende Regelkreis (daher kann Messung 11 ausgeschlossen werden) wird auch in diesem Fall korrekt erkannt.

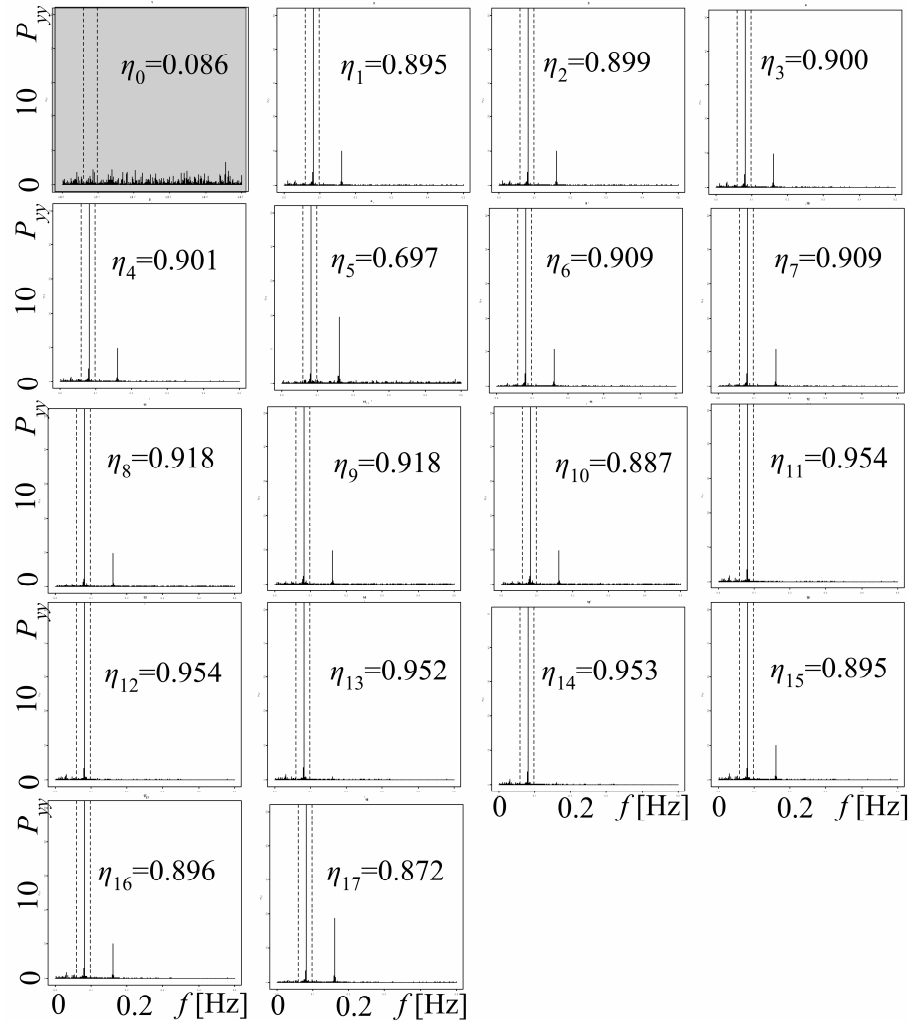


Abbildung 5.9: Spektren aller Teilprozesse bei Verursacher S_{11} (grau unterlegt: ausgeschlossen von Analyse, da dort die anlagenweite Störungsfrequenz nicht auffindbar ist)

Detaillierte Fehlerdiagnose im verursachenden Teilprozess S_{11}

Basierend auf den Ergebnissen der erfolgreichen Verursacheridentifikation beider Verfahren und der in Abbildung 2.2 vorgeschlagenen Prozedur muss im nächsten Schritt eine Fehlerdiagnose im Teilprozess S_{11} durchgeführt werden. Dazu werden die in Kapitel 4 vorgestellten Verfahren verwendet. Es werden die Performance-Indizes gemäß Tabelle 5.2 berechnet. Es fällt auf, dass die Performance-Indizes den schwingenden Sollwert korrekt erkennen und gleichzeitig erhöhtes Messrauschen diagnostizieren.

Tabelle 5.2: Berechnete Performance-Indizes des Benchmarkprozesses bei schwingendem Sollwert in S_{11}
(grau hinterlegt: Anforderung erfüllt, schraffiert: Anforderung nicht erfüllt)

	p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8	p_9
S_{11}	0.028 ↓	0 ↓	0.208 ↓	0.15 ↑	0 ↓	0.045 ↓	0.932 ↑	0.932 ↑	0.012 ↓
Aggressive Reglereinstellung	↓	→	→	→	↓	↓	↑	↓	→
Schwingender Sollwert	→	→	→	→	→	→	→	↑	→
Messrauschen	→	→	→	↑	→	→	→	→	→
Begrenzungen	↑	↑	↑	→	→	→	→	→	→
Inaktive Regelung	↑	↑	↓	→	→	→	→	→	→
Haft- und Gleitreibung	→	→	→	→	↑	↑	↑	↓	→
Externe Störungen	→	→	→	→	→	→	→	→	↑

5.2 Industrielle Benchmarks

In Abschnitt 5.1 wurde gezeigt, dass das vorgestellte Verfahren zur anlagenweiten Regelkreisdiagnostik für einen simulierten Benchmark einwandfrei funktioniert. Im Folgenden soll dasselbe Verfahren genutzt werden, das Verhalten eines industriellen Hydrocrackers zu diagnostizieren. In unregelmäßigen Abständen weisen nahezu alle Messungen charakteristische Schwingungen auf. In Abbildung 5.10 sind die industriellen Daten für einen Zeitraum von 500 min gezeigt. Die Messstellenbezeichnungen richten sich nach dem in Abbildung 5.1 gezeigten Fließbild.

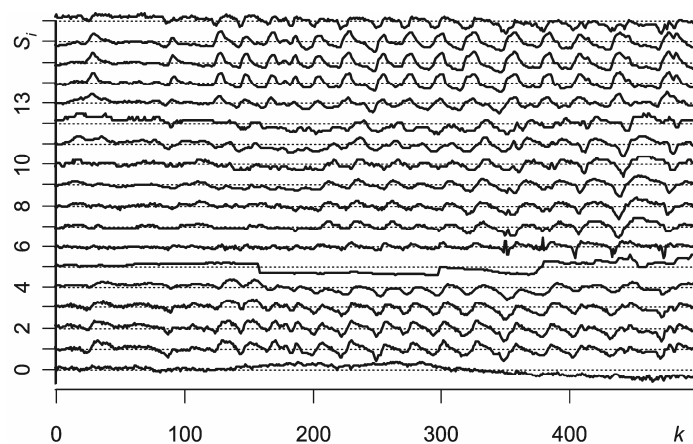


Abbildung 5.10: Zeitlicher Verlauf aller relevanten Hydrocrackermessungen [74]

Die ähnlichen Schwingungsverläufe deuten nach [60] eindeutig auf eine anlagenweite Störung hin. Daher muss zunächst der verursachende Teilprozess

identifiziert werden und anschließend gemäß der Prozedur aus Abbildung 2.2 verfahren werden. Soll dies mit minimalem Rechenaufwand geschehen, so ist das in Kapitel 2.3 gezeigte Verfahren zu verwenden. Steht mehr Rechenleistung und Rechenzeit zur Verfügung, so sei das Verfahren aus Abschnitt 2.2 zu verwenden. Bei der vorliegenden Validierung sollen beide Ansätze verwendet werden. Zunächst die rechenintensivere Methode, basierend auf der k -nächsten-Nachbarn-Ersetzung.

Offline-Verursacheridentifizierung mittels KNNK Algorithmus gemäß Abschnitt 2.2

Basierend auf der in Abbildung 2.2 gezeigten Prozedur, muss zunächst die Beeinflussungsmatrix bestimmt werden. Da es sich bei Messung 1 um die externe Störung S_0 handelt, werden diese Daten von der Berechnung ausgeschlossen. Für den Datensatz aus Abbildung 5.10 ergibt sich die Beeinflussungsmatrix gemäß

$$\mathbf{Z} = \begin{bmatrix} - & & & & & & & & & & & & & & & & \\ 1 & - & & & & & & & & & & & & & & & \\ 1 & 2 & - & & & & & & & & & & & & & & \\ 1 & 2 & 3 & - & & & & & & & & & & & & & \\ 1 & 2 & 5 & 5 & - & & & & & & & & & & & \\ 1 & 6 & 3 & 4 & 6 & - & & & & & & & & & & \\ 1 & 2 & 3 & 4 & 5 & 6 & - & & & & & & & & & \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & - & & & & & & & & \\ 1 & 9 & 3 & 4 & 9 & 6 & 7 & 8 & - & & & & & & & \\ 1 & 10 & 10 & 10 & 10 & 6 & 7 & 8 & 9 & - & & & & & & \\ 1 & 11 & 11 & 11 & 11 & 6 & 7 & 8 & 9 & 11 & - & & & & & \\ 12 & 12 & 12 & 12 & 12 & 6 & 7 & 8 & 9 & 10 & 11 & - & & & & \\ 13 & 13 & 13 & 13 & 13 & 13 & 13 & 13 & 13 & 13 & 13 & 13 & - & & & \\ 14 & 14 & 14 & 14 & 14 & 14 & 7 & 8 & 9 & 10 & 14 & 12 & 13 & - & & \\ 15 & 15 & 15 & 15 & 15 & 15 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & - & \\ 16 & 16 & 16 & 16 & 16 & 16 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 & - \\ 1 & 2 & 17 & 17 & 17 & 17 & 17 & 17 & 17 & 10 & 11 & 17 & 13 & 17 & 17 & 17 & - \end{bmatrix}. \quad (5.3)$$

Die Beeinflussungsmatrix identifiziert Messung 13 als Verursacher. Da es sich hierbei nicht um den Teilprozess (S_1) handelt, welcher der äußeren Störung am nächsten liegt, muss im nächsten Schritt eine detaillierte Fehlerdiagnose in diesem Teilprozess durchgeführt werden.

Online-Verursacheridentifizierung anhand der Störungsenergie gemäß Abschnitt 2.3

Die vereinfachte Verursacheridentifikation sei an dieser Stelle nur verwendet um die bereits genannten Einschränkungen des Verfahrens zu verdeutlichen. Der Grund für die vermutete Unwirksamkeit liegt darin, dass die Anlage nicht durch eine einzelne harmonische, sinusförmige Oszillation gestört wird. Die Spektren der einzelnen Messungen ergeben sich gemäß den Darstellungen in Abbildung 5.11.

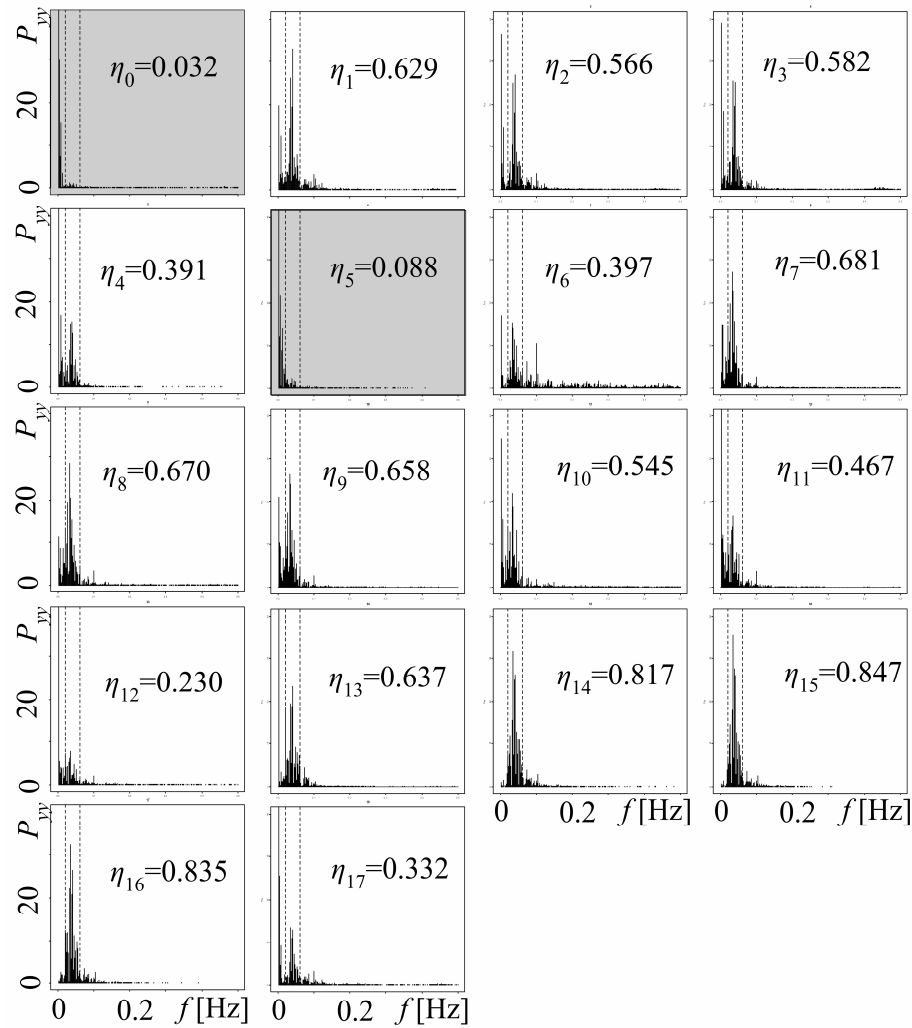


Abbildung 5.11: Spektren aller Teilprozesse bei unbekanntem Verursacher (grau unterlegt: ausgeschlossen von Analyse, da dort die anlagenweite Störungsfrequenz nicht auffindbar ist)

Werden sowohl die Messungen 0 als auch 5 vernachlässigt, so lässt sich die größte anlagenweite Störung bei einer Frequenz von 0,04 Hz zu finden und das Störungsfrequenzband ist von 0,02 bis 0,06 Hz zu finden. In Abbildung 5.11 ist dieser Bereich als senkrechte Linie gekennzeichnet. Die daraus berechneten relativen Störungsenergien (η_0 bis η_{17}) sind ebenfalls dargestellt.

Ein Blick auf die berechneten Werte zeigt, dass die zu Beginn geäußerte Vermutung sich bestätigt und Messung 15 fälschlicherweise als Verursacher erkannt wird. Der Grund für diese Fehleinschätzung liegt darin, dass der Verlauf von Messung 15 am ehesten einer sinusförmigen Schwingung ähnelt. Diese Tatsache kann eventuell dadurch erklärt werden, dass der zur Messung 15 gehörende Regelkreis zu aggressiv versucht, die von Messung 13 hervorgerufenen Störungen auszugleichen.

Detaillierte Fehlerdiagnose im verursachenden Teilprozess S_{13}

Basierend auf den Ergebnissen der ausführlichen Verursacheridentifikation und der in Abbildung 2.2 vorgeschlagenen Prozedur soll im Folgenden eine ausführliche Fehlerdiagnose im verursachenden Teilprozess S_{13} durchgeführt werden. Dazu werden die in Kapitel 4 vorgestellten Verfahren verwendet. Es werden die Performance-Indizes gemäß Tabelle 5.3 berechnet.

Tabelle 5.3: Berechnete Performance-Indizes des industriellen Prozesses
(grau hinterlegt: Anforderung erfüllt, schraffiert: Anforderung nicht erfüllt)

	p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8	p_9
S_{14}	0.022 ↓	0 ↓	0.038 ↓	0.032 ↓	0 ↓	0.350 ↑	0.827 ↑	0.772 ↑	0.002 ↓
Aggressive Reglereinstellung	↓	→	→	→	↓	↓	↑	↓	→
Schwingender Sollwert	→	→	→	→	→	→	→	↑	→
Messrauschen	→	→	→	↑	→	→	→	→	→
Begrenzungen	↑	↑	↑	→	→	→	→	→	→
Inaktive Regelung	↑	↑	↓	→	→	→	→	→	→
Haft- und Gleitreibung	→	→	→	→	↑	↑	↑	↓	→
Externe Störungen	→	→	→	→	→	→	→	→	↑

Mithilfe der Berechnung ist es möglich das Regelungsszenario des schwingenden Sollwerts zu diagnostizieren. Dass diese Überlegung korrekt ist, zeigt die Darstellung in Abbildung 5.12. Der diagnostizierte schwingende Sollwert ergab sich aus der überlagerten Kaskadenregelung. Als diese daraufhin abgeschaltet wurde und die Anlage im Handbetrieb gefahren wurde, klang die Oszillation nach einiger Zeit ab.

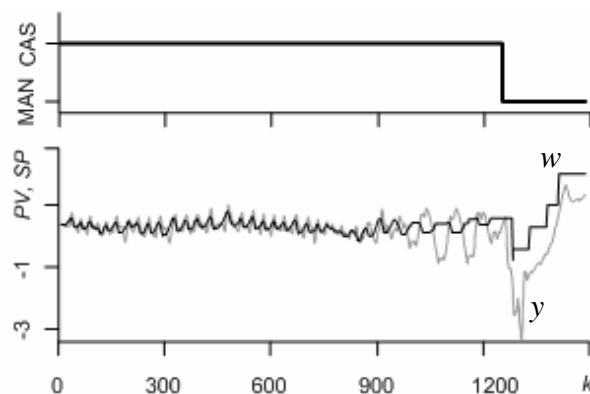


Abbildung 5.12: Wechsel von Kaskadenregelung (CAS) auf manuelle Regelung (MAN) führt nach Diagnose zum Abklingen der Schwingung

Kapitel 6

Zusammenfassung

Die vorliegende Arbeit beschreibt ein neues Verfahren zur anlagenweiten Regelkreisdiagnostik. Der Passus „anlagenweit“ beschreibt, dass eine gesamte Anlage, bestehend aus einer Vielzahl von Regelkreisen gleichzeitig auf ihre Regelungsperformance untersucht wird. Bei der Entwicklung desselben wurde dabei besonderer Wert auf die industriellen Anforderungen an ein solches Verfahren gelegt, was dazu führt, dass sämtliche Algorithmen online sogar im Prozessleitsystem implementierbar sind und der theoretische Hintergrund für Jedermann nachvollziehbar ist, ohne große Einbußen bei der Effektivität mit sich zu ziehen. Das entwickelte Verfahren zur anlagenweiten Regelkreisdiagnostik besteht dabei aus den Schritten der Verursacheridentifikation und der Regelkreisdiagnose.

Das Ziel des ersten Schrittes, der Identifikation, ist es aus sämtlichen Regelkreisen einer Anlage gegebenenfalls jenen zu identifizieren, der die restlichen Regelkreise hauptsächlich beeinflusst. Im zweiten Schritt, der Diagnose, wird beurteilt, ob ein Fehler in diesem Regelkreis der Grund für diese signifikante Beeinflussung ist. Ein weiteres Ergebnis der Diagnose ist das Auffinden der Ursache für diese Art der anlagenweiten Störung, so dass entweder direkt Verbesserungen (z.B. Reglerneueinstellung) vorgenommen werden oder aber zumindest geplant werden können (z.B. im Falle von Haft- und Gleitreibung). Die Verursacheridentifikation wird mithilfe eines neu entwickelten Verfahrens der nichtlinearen Totzeitschätzung realisiert. Diese Art der Berechnung ermöglicht eine Analyse ohne jegliche Prozesskenntnis oder Modellannahme, welche dabei ebenfalls vollständig automatisierbar ist. Der Algorithmus und die daraus abgeleiteten Ergebnisse werden in der vorliegenden Arbeit ebenfalls theoretisch validiert, sowohl für statische und dynamische als auch für Systeme mit mehreren Eingängen. Basierend auf der Tatsache, dass dieses neue Verfahren der Verursacheridentifikation aufgrund seiner vielen Rechenschritte und verschachtelten Schleifen sehr rechenintensiv ist, wurde ein weiteres vereinfachtes Verfahren vorgestellt, welches auch bei kürzeren Abtastzeiten online implementiert ist. Der damit verbundene Nachteil wurde in der vorliegenden Arbeit herausgearbeitet. Er besteht darin, dass dieses Verfahren am verlässlichsten bei der Erkennung von Verursachern von sinusförmigen Schwingungen arbeitet, wie es etwa bei aggressiven Reglern, schwingenden Sollwerten oder sinusförmigen externen Störungen der Fall ist. Es wurde in der vorliegenden Arbeit gezeigt, dass das komplexere Verfahren diese Einschränkungen nicht besitzt, da es auch bei nicht periodischen und sogar stochastischen Störungen zuverlässig arbeitet.

Zur Realisierung des zweiten Schrittes, der Regelkreisdiagnostik wurden zehn Indizes eingeführt, die ihrerseits nahezu einfach zu programmieren sind und entweder selbst entwickelt oder der genannten Literatur entnommen sind. Zu den selbst entwickelten Kennzahlen zählen dabei:

- relative Zeit des stationären und signifikanten Regelfehlers,
- relative Rauschintensität,
- Gleichverteilung des Stellsignals,
- normalisierte Oszillationsindex für Sollwert und Regelgröße,
- relative Beeinflussung durch deterministische externe Störungen.

Es wurde gezeigt, dass verschiedene Regelkreisprobleme spezifische Konstellationen der zehn Indizes hervorrufen, was wiederum dazu führt, dass jedes Problem eindeutig diagnostiziert werden kann.

Das Konzept der anlagenweiten Regelkreisdiagnostik wurde in der vorliegenden Arbeit ebenfalls an zahlreichen kleineren und verschiedenen komplexen Simulationen erfolgreich validiert. Basierend auf dieser Validierung wurde das Verfahren ebenfalls an einer bestehenden Industrieanlage getestet, wobei der Grund einer anlagenweiten Störung erfolgreich aufgedeckt werden konnte. Die Erkenntnisse dieser Diagnose konnten zeitnah verarbeitet und umgesetzt werden, was zu einer schnellen und signifikanten Verbesserung der Anlagen-Performance führte.

Literaturverzeichnis

- [1] K. J. Aström. Computer control of a paper machine – an application of linear stochastic control theory. *IBM J. Res. Dev.*, 11, 389-405, 1967.
- [2] M. Bauer, J. W. Cox, M. H. Caveness, J. J. Downs, N. F. Thornhill. Finding the direction of disturbance propagation in a chemical process using transfer entropy, *IEEE Transactions on Control System Technology*, 15, 12-21, 2007.
- [3] G. E. P. Box, G. M. Jenkins. Time series analysis: forecasting and control, Oakland, CA: *Holden-Day*. 1976.
- [4] S. Cao, R. R. Rhinehart. An efficient method for on-line identification of steady-state, *Journal of Process Control*, 5, 363-374, 1995.
- [5] M. A. A. S. Choudhury, N. F. Thornhill, S. L. Shah. Modelling Valve Stiction. *Control Engineering Practice*, 13, 641-658, 2005.
- [6] M.A.A. S. Choudhury, S. L. Shah, N. F. Thornhill. Detection and diagnosis of system nonlinearities using higher order statistics. *Automatica*, 40, 1719-1728, 2004.
- [7] T. M. Cover, P. E. Hart. Nearest neighbor pattern classification. *IEEE Trans. Inform. Theory*, IT-13(1), 21-27, 1967.
- [8] A. Davis. Small sample effects on information-theoretic estimates, in: 43rd Annual Meeting of the APS Division of Plasma Physics, California, USA, 2001.
- [9] L. Desborough, P. Nordh, R. Miller. Control system reliability: process out of control. *Ind Comput*, 8, 52-55, 2001.
- [10] L. Desborough, R. Miller. Increasing customer value of industrial control performance monitoring - Honeywell's experience. *AIChE Symp. Ser. No 326*, 98, 153-186, 2002.
- [11] L. Desborough, T. Harris, Performance assessment measures for univariate feedback control, *Can. J. Chem. Eng.*, 70, 1186-1197, 1992.
- [12] R. Dittmar, B.-M. Pfeiffer. *Modellbasierte prädiktive Regelung*, Oldenbourg Verlag, 2004.
- [13] G. A. Dumont, L. Kammer, B. J. Allison, L. Ettaleb, A. A. Roche. Control performance monitoring: New developments and practical issues. In *Proceedings of the IFAC world congress*, Barcelona, Spain, 2002.
- [14] EEMUA 191 Alarm Systems - A Guide to Design, Management and Procurement, ISBN 0-85931-076-0, 1999.

- [15] D. B. Ender. Process Control Performance: Not as good as you think. *Control Eng.*, 9, 180-190, 1993.
- [16] L. Fahrmeir, R. Künstler, I. Pigeot, G. Tutz. *Statistik*, Springer Verlag, Berlin, Germany, 2002.
- [17] B. Fitzgerald. Control Valve for the Chemical Process Industries, McGraw-Hill, New York, 1995.
- [18] K. Forsmann. On detection and classification of valve stiction. in *Proceedings of TAPPI Conf Process Control*, Williamsburg, USA, 2000.
- [19] K. F. Früh, U. Maier, D. Schaudel. *Handbuch der Prozessautomatisierung*, Oldenbourg Industrieverlag, S. 147, 2008.
- [20] E. Gandert. Schaurige Schauer - Alarmmanagement in der Prozessindustrie, Chemie Technik, *Kompodium Prozessautomation*, 36-39, 2009.
- [21] K. M. Grigoriadis, I. F. Fialho, F. Zhang. Linear Parameter-Varying Anti-windup Control For Active Microgravity Isolation, *ISSO Annual Report Y2003*, University of Houston, Clear Lake, 52-57, 2003.
- [22] R. Haber, L. Keviczky. *Nonlinear system identification*, Kluwer Academic Publishers, MA, USA, 1999.
- [23] M. Hadjiiski, Z. Georgiev. Benchmarking of Process Control Performance, University of Chemical Technology and Metallurgy, Sofia.
<http://www.iit.bas.bg/PECR/55/103-110.pdf>, nt, 14.11.2011.
- [24] T. Hägglund. Automatic detection of sluggish control loops. *Control Engineering Practice*, 7(12), 1505-1511, 1999.
- [25] T. Harris. Assessment of control loop performance. *The Canadian Journal of Chemical Engineering*, 67, 856-861, 1989.
- [26] Q. P. He, M. Pottmann, Detection of valve stiction using curve fitting. Internal Report, *Process Dynamics and Control*, DuPont Engineering, 2003.
- [27] A. Horch, A. J. Isaksson. Detection of valve stiction in integrating processes. Internal Report, Royal Institute of Technology, Stockholm, Sweden, IR-S3-REG-0006, 2000.
- [28] A. Horch. A simple method for detection of stiction in control valves. *Control Engineering Practice*, 7, 1221-1231, 1999.
- [29] A. Horch. Asset Productivity – Innovative Betriebsführung von Industrieanlagen und Kraftwerken, in: *Pressetag im ABB Forschungszentrum*, Ladenburg, Deutschland, 2004.
- [30] A. Horch. Condition monitoring of control loops. *PhD Thesis*, Royal Institute of Technology, Sweden, 2000.
- [31] A. Horch. Method and a system for Evaluation of Static Friction. *US Patent 7047152*, 2006.
- [32] R. Howard, D. Cooper. A novel pattern-based approach for diagnostic controller performance monitoring. *Control Engineering Practice*, 18, 3, 279-288, 2010.

- [33] B. Huang. Minimum variance control and performance assessment of timevariant processes. *Journal of Process Control*, 12, 707-719, 2002.
- [34] A. Ingimundarson. Dead-time compensation and performance monitoring in process control, *PhD thesis*, Lund Institute of Technology, Sweden, 2003.
- [35] G. Jacovitti, G. Scarano. Discrete time techniques for time delay estimation, *Signal Processing*, 41, 525-533, 1993.
- [36] S.-L. Jämsä-Jounela, R. Poikonen, N. Vatanski, A. Rantala. Evaluation of control performance: methods and applications in a flotation plant. *Minerals Engineering*, 16, 1069-1074, 2003.
- [37] M. Jelali, An overview of control performance assessment technology and industrial applications, *Control Engineering Practice*, 14, 441-466, 2006.
- [38] M. Jelali, B. Huang. *Detection and Diagnosis of Stiction in Control loops*, Springer Verlag, London, 2010.
- [39] M. Kano, H. Maruta, H. Kugemoto, K. Shimizu. Practical model and detection algorithm for valve stiction. *Proc. IFAC Dycops*, Cambridge, USA, 2004.
- [40] A. Kayihan, F. J. Doyle. Friction compensation for a process control valve, *Contr. Eng. Pract.*, 799-812, 2000.
- [41] B. Ko, T. F. Edgar. Assessment of achievable PI controller performance for linear process with dead time. In *American Control Conference*, Philadelphia, USA, 1998.
- [42] D. Kozub, Controller performance monitoring and diagnosis, *Proceedings of Fifth International Conference on Chemical Process Control*, CACHE Corp, 83-96, 1997.
- [43] P. Langley, W. Iba, and K. Thompson. An analysis of Bayesian classifiers. In *Proceedings of the Tenth National Conference on Artificial Intelligence*, San Jose, CA, 399-406., 1992.
- [44] J. Lunze, *Automatisierungstechnik*, Oldenbourg, München, 2007.
- [45] Namur-Arbeitskreis 2.9, Alarm Management - NA 102, 2003, zu beziehen unter <http://www.namur.de/arbeitsfelder-af-und-projektgruppen-pg/arbeitsfeld-2-prozess-und-betriebsfuehrungssysteme/ak-29-mensch-prozess-kommunikation/?L=0>, 28.03.2012.
- [46] nt, http://www.chemiedidaktik.uni-wuppertal.de/alte_seite_du/material/exarbeiten/pmma/pmma2.pdf, Stand 27.02.2012.
- [47] A. O'Dwyer. *Handbook of PI and PID controller tuning rules*, Imperial College Press, London, 2006.
- [48] A. Ordys, D. Udachi, M. A. Johnson. *Process Control performance Assessment – from theory to implementation*, Springer, London, 227-257, 2007.
- [49] S. J. Qin. Control Performance Monitoring – a review and assessment, *Computers and Chemical Engineering*, 23, 173-186, 1998.

- [50] T. I. Salsbury, A. Singhal, Shape-based Stiction detection using area Calculations, in M. Jelali, B. Huang. *Detection and Diagnosis of Stiction in Control loops*, Springer Verlag, London, 183-204, 2010.
- [51] C. Scali, C. Ghelardoni. An improved qualitative shape analysis technique for automatic detection of valve stiction in flow control loops. *Contr Eng Pract*, 16, 1501-1508, 2008.
- [52] T. Schreiber. Measuring Information Transfer, *Physical review letters - Max Planck Institute for the Physics of Complex Systems*, Dresden, Germany, 85, 461-464, 2000.
- [53] C. E. Shannon. Prediction and entropy of printed English, *The Bell System Technical Journal*, 30, 50-64, 1951.
- [54] A. Singhal, T. Salsbury. A simple method for detecting valve stiction in oscillating control loops. *Journal of Process Control*, 15, 371-382, 2005.
- [55] R. Srinivasan, R. Rengaswamy, R. Miller. Control loop performance assessment. 1. A qualitative approach for stiction diagnosis. *Ind Eng Chem Res*, 44, 6708-6718, 2005.
- [56] I. Steinwart, A. Christmann. *Support Vector Machines*, Springer Verlag, London, 2008.
- [57] N. F. Thornhill, B. Huang, H. Zhang. Detection of multiple oscillations in control loops. *Journal of Process Control*, 13, 91-100, 2003.
- [58] N. F. Thornhill, M. A. A. S. Choudhury, S. L. Shah. The impact of compression on data-driven process analyses. *Journal of Process Control*, 14, 389-398, 2004.
- [59] N. F. Thornhill, M. Oettinger, P. Fedenczuk. Refinery-wide control loop performance assessment. *Journal of Process Control*, 9, 109-124, 1999.
- [60] N. F. Thornhill, S. L. Shal, B. Huang. Detection of distributed oscillation and root-cause diagnosis. in *Proceedings of CHEMFAS 4*, Korea, 167-172, 2001.
- [61] A. J. Trenchard. Loop Management for the typical process user, in: *IEE Seminar on Control Loop Assessment and Diagnosis*, London, UK, 2005.
- [62] O. Troyanskaya, M. Cantor, G. Sherlock, P. Brown, T. Hastie, R. Tibshirani, D. Botstein, R. B. Altman. Missing value estimation methods for DNA microarrays, *Bioinformatics*, 17(6), 520-525, 2001.
- [63] H. Unbehauen. *Regelungstechnik III - Identifikation, Adaption, Optimierung*. Vieweg Verlag, 2000.
- [64] R. Vaught, J. Tippet. Control performance monitoring: Shaman or saviour. *Pulp & Paper*, 102, 26-29, 2001.
- [65] B. R. Whalen. *Basic Instrumentation, petroleum extension service*, Austin, Texas, 1983.
- [66] C. Xia , J. Howell. Isolating multiple sources of plant-wide oscillations via independent component analysis, *Control Engineering Practice*, 13, 1027-1035, 2005.

- [67] M. Yuwana, D. E. Seborg, A new method for on-line controller tuning, *AIChE Journal*, 28, 434-440, 1982.
- [68] X. Zang, J. Howell. Isolating the source of whole-plant oscillations through bi-amplitude ratio analysis, *Control Engineering Practice*, 15, 69-76, 2007.
- [69] H. Zhang, Y. Fu, Z. Tang. Root-Cause Analysis based on time delay estimation for alarm systems in power grids, *Advanced Materials Research*, 344, 943-946, 2011.

Beiträge für Zeitschriften und Konferenzen, die im Rahmen dieser Arbeit entstanden sind

- [70] M. Stockmann, R. Haber. Determination of fault propagation by time delay estimation using k nearest neighbour imputation, in *Proceedings of Sys'tol 2010 - Conference on Control and Fault Tolerant Systems*, France, Nice, 2010.
- [71] M. Stockmann, R. Haber. Extensive control performance monitoring using the limited potential of today's process control systems. in *Proceedings of 9th European Workshop on Advanced Control and Diagnosis*, Budapest, Hungary, 2011.
- [72] M. Stockmann, R. Haber, U. Schmitz. Identification of plant-wide oscillation source by disturbing energy transfer determination, in *Proceedings of 11th International Carpathian Control Conference*. Eger, Hungary, 2010.
- [73] M. Stockmann, R. Haber, U. Schmitz. Practical Application of a New Time Delay Estimation Method for Control Performance Monitoring. *ACD 2010 - 8th European Workshop on Advanced Control and Diagnosis*, Ferrara, Italy, 2010.
- [74] M. Stockmann, R. Haber, U. Schmitz, Source identification of plant-wide faults based on a k nearest neighbor time delay estimation. *Journal of Process Control*, 22 (3), 583-598. 2012.
- [75] M. Stockmann, R. Haber, U. Schmitz, Valve Stiction Detection by Pattern Recognition and Principal Component Analysis, in: *Proceedings of 7th IFAC International Symposium on Fault Detection, Supervision and Safety of Technical Processes*, Spain, Barcelona, 2009.